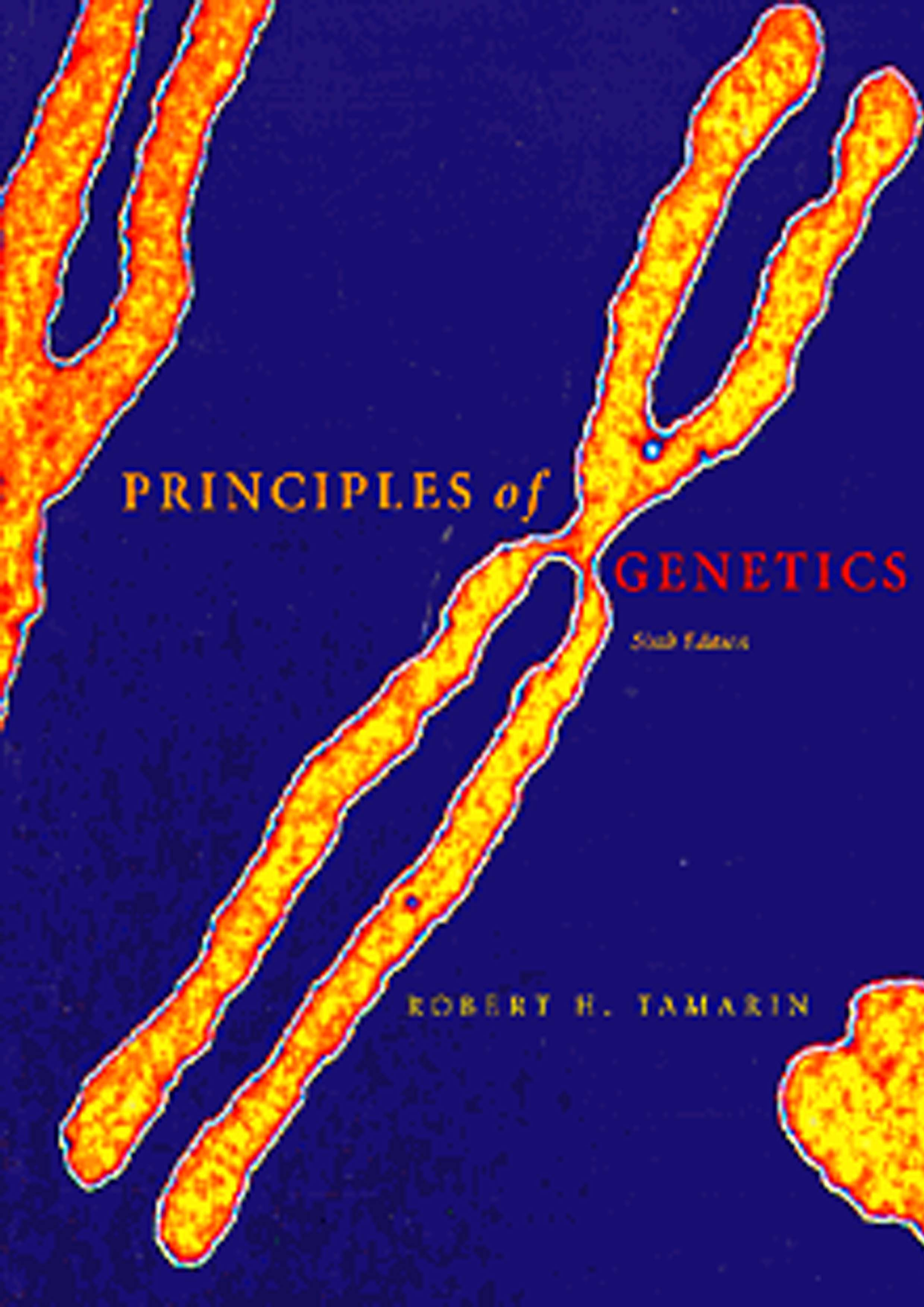


# PRINCIPLES of GENETICS

Sixth Edition

ROBERT H. TAMARIN

VISIT…

LANZAROTE

Caliente.COM

# PREFACE

The twentieth century began with the rediscovery of Mendel's rules of inheritance and ended with the complete sequence of the human genome, one of the most monumental scientific accomplishments of all time. What lies in the future? What will the twenty-first century, the century of genomics, bring? Will geneticists a hundred years from now speak of a complete cure for cancer, heart disease, and mental illness? Will we have a cure for autoimmune diseases such as diabetes and arthritis? Will aging be slowed or even prevented? Will we have a complete understanding of the process of development and a concurrent elimination of birth defects and developmental problems? Will genetics put an end to world hunger? How will we live, and what will be the quality of our lives? The students who now are taking genetics will learn the answers to these questions as time progresses. Some students will contribute to the answers.

The science of genetics includes the rules of inheritance in cells, individuals, and populations and the molecular mechanisms by which genes control the growth, development, and appearance of an organism. No area of biology can truly be appreciated or understood without an understanding of genetics because genes not only control cellular processes, they also determine the course of evolution. Genetic concepts provide the framework for the study of modern biology.

This text provides a **balanced treatment** of the major areas of genetics in order to prepare the student for upper-level courses and to help share in the excitement of research. Most readers of this text will have taken a general biology course and will have had some background in cell biology and organic chemistry. For an understanding of the concepts in this text, however, the motivated student will need to have completed only an introductory biology course and have had some chemistry and algebra in high school.

Genetics is commonly divided into three areas: classical, molecular, and population, although molecular advancements have blurred these distinctions. Many genetics teachers feel that a historical approach provides a sound introduction to the field and that a thorough grounding in Mendelian genetics is necessary for an understanding of molecular and population genetics—an approach this text follows. Other teachers, however, may prefer to begin with molecular genetics. For this reason, the chapters have been grouped as **units that allow for flexibility** in their use. A comprehensive glossary and index will help maintain continuity if the instructor chooses to change the order of the chapters from the original.

An understanding of genetics is crucial to advancements in medicine, agriculture, and many industries. Genetic controversies—such as the pros and cons of the Human Genome Project, the potential ethical and medical risks of recombinant DNA and cloning of mammals, and human behavioral genetic issues such as the degree of inheritance of homosexuality, alcoholism, and intelligence—have captured the interest of the general public. Throughout this text, we examine the **implications for human health and welfare** of the research conducted in universities and research laboratories around the world; boxed material in the text gives insight into genetic techniques, controversies, and breakthroughs.

Because genetics is the first analytical biology course for many students, some may have difficulty with its quantitative aspects. There is no substitute for work with pad and pencil. This text provides a larger number of problems to help the student learn and retain the material. All problems within the body of the text and a selection at the end of the chapters should be worked through as they are encountered. After the student has worked out the problems, he or she can refer to the answer section in Appendix A. We provide solved problems at the end of each chapter to help.

In this text, we stress **critical thinking,** an approach that emphasizes understanding over memorization, experimental proof over the pronouncements of authorities, problem solving over passive reading, and active participation in lectures. The latter is best accomplished if the student reads the appropriate text chapter before coming to lecture rather than after. That way the student can use the lecture to gain insight into difficult material rather than spending the lecture hectically transcribing the lecturer's comments onto the notebook page.

For those students who wish to pursue particular topics, a **reference section** in the back of the text provides chapter-by-chapter listings of review articles and articles in the original literature. Although some of these articles might be difficult for the beginner to follow, each is a landmark paper, a comprehensive summary, or a paper with some valuable aspect. Some papers may contain an insightful photograph or diagram. Some magazines and journals are especially recommended for the student to look at periodically, including *Scientific American,* *Science,* and *Nature,* because they contain nontechnical summaries as well as material at the cutting edge of genetics. Some articles are included to help the instructor find supplementary materials related to the concepts in this book. Photographs of selected geneticists also are included. Perhaps the glimpse of a face from time to time will help add a human touch to this science.

The World Wide Web also can provide a valuable resource. The textbook has its own website: **www.mhhe.com/tamarin7**. In addition, the student can find much material of a supplemental nature by "surfing" the web. Begin with a search engine such as: www.yahoo.com, or www.google.com and type in a key word. Follow the links from there. Remember that the material on the web is "as is"; it includes a lot of misinformation. Usually, content from academic, industrial, and organizational sources is relatively reliable; however, caveat emptor—buyer beware. Often in surfing for scientific key words, the student will end up at a scientific journal or book that does not have free access. Check with the university librarian to see if access might be offered to that journal or book. The amount of information that is accurate and free is enormous. Be sure to budget the amount of time spent on the Internet.

## NEW TO THIS EDITION

Since the last edition of this text, many exciting discoveries have been made in genetics. All chapters have been updated to reflect those discoveries. In particular:

- The chapter on Recombinant DNA Technology has been revised to be a chapter on Genomics, Biotechnology, and Recombinant DNA (sixth edition chapter 12 has become chapter 13 in this edition). The chapter includes new material on the completion of the Human Genome Project, bioinformatics, proteomics, and the latest techniques in creating cDNA and knockout mice.
- The chapter on Control of Transcription in Eukaryotes (sixth edition chapter 15 has become chapter 16 in this edition) has been completely reorganized and rewritten to emphasize signal transduction, specific transcription factors, methylation, and chromatin remodeling in control of gene expression; as in the last edition, there are specific sections on *Drosophila* and plant development, cancer, and immunogenetics.
- For better continuity, the chapter on Mutation, Recombination, and DNA Repair has been moved to follow the chapters on Transcription and Translation (sixth edition chapter 16 has become chapter 12 in this edition).
- The material in chapter 3 on Genetic Control of the Cell Cycle has been upgraded to a chapter section on the Cell Cycle.
- Molecular material throughout the book has been completely updated to include such subjects as numerous DNA repair polymerases and their functioning; base-flipping; TRAP control of attenuation; and chromatosomes.

## LEARNING AIDS FOR THE STUDENT

To help the student learn genetics, as well as enjoy the material, we have made every effort to provide pedagogical aids. These aids are designed to help organize the material and make it understandable to students.

- **Study Objectives** Each chapter begins with a set of clearly defined, page-referenced objectives. These objectives preview the chapter and highlight the most important concepts.
- **Study Outline** The chapter topics are provided in an outline list. These headings consist of words or phrases that clearly define what the various sections of the chapter contain.
- **Boldface Terms** Throughout the chapter, all new terms are presented in boldface, indicating that each is defined in the glossary at the end of the book.
- **Boxed Material** In most chapters, short topics have been set aside in boxed readings, outside the main body of the chapter. These boxes fall into four categories: **Historical Perspectives, Experimental Methods, Biomedical Applications,** and **Ethics and Genetics.** The boxed material is designed to supplement each chapter with entertaining, interesting, and relevant topics.
- **Full Color Art and Graphics** Many genetic concepts are made much clearer with full-color illustrations and the latest in molecular computer models to help the student visualize and interpret difficult concepts. We've added thirty new photographs and over a hundred new and modified line drawings to this edition.
- **Summary** Each chapter summary recaps the study objectives at the beginning of the chapter. Thus, the student can determine if he or she has gained an understanding of the material presented in the study objectives and reinforce them with the summary.
- **Solved Problems** From two to four problems are worked out at the end of each chapter to give the student practice in solving and understanding basic problems related to the material.
- **Exercises and Problems** At the end of the chapter are numerous problems to test the student's

understanding of the material. These problems are grouped according to the sections of the chapter. Answers to the odd-numbered problems are presented in Appendix A, with the even-numbered problems answered only in the Student Study Guide so that the student and instructor can be certain that the student is gaining an understanding of the material.

- **Critical Thinking Questions** Two critical thinking questions at the end of each chapter are designed to help the student develop an ability to evaluate and solve problems. The answer to the first critical thinking question can be found in Appendix A, and the answer to the second question is in the Student Study Guide.

## ANCILLARY MATERIALS

### For the Instructor

- *Website.* Visit us at **www.mhhe.com/tamarin7.** Here instructors will find jpeg files of the line drawings and tables suitable for downloading into PowerPoint, quizzes for study support, and links to genetic sites. In addition, instructors will also find a link to our hugely successful **PageOut: The Course Website Development Center,** where instructors can create a professional-looking, customized course website. It's incredibly easy to use, and you need not know html coding.
- *Visual Resource Library (VRL).* This Windows- and Macintosh-compatible CD-ROM has all the line drawings and tables from the text suitable for PowerPoint presentations. (ISBN 0072334266)
- *Instructor's Manual with Test Item File.* Available on the website, the Instructor's Manual contains outlines, key words, summaries, instructional hints, and supplemental aids. The Test Item File contains 35 to 50 objective questions with answers for each chapter. (ISBN 0072334215)
- *Test Item File on MicroTest III Classroom Testing Software* is an easy-to-use CD-ROM test generator also offered free upon request to adopters of this text. The software requires no programming experience and is compatible with Windows or Macintosh systems. (ISBN 0072334231).

### For the Student

- *Website.* Visit us at **www.mhhe.com/tamarin7.** Here the student will find quizzes for study support, web exercises and resources, and links to genetic sites.
- *Genetics: From Genes to Genomes CD-ROM,* by Ann E. Reynolds, University of Washington. Packaged free with every text, this CD-ROM covers the most challenging concepts in the course and makes them more understandable through the presentation of full-color, narrated animations and interactive exercises. The text indicates related topics on the CD with the following icon: 💿
- *Student Study Guide.* This study guide features key concepts, problem-solving hints, practice problems, terms, study questions, and answers to even-numbered questions in the text. (ISBN 0072334207)
- *Laboratory Manual of Genetics 4/e,* by A. M. Winchester and P. J. Wejksnora, University of Wisconsin–Milwaukee. This manual for the genetics laboratory features classical and molecular biology exercises that give students the opportunity to apply the scientific method to "real"—not simulated—lab investigations. (ISBN 0697122875)
- *Case Workbook in Human Genetics, 2/e,* by Ricki Lewis, SUNY–Albany. The Workbook includes thought-provoking case studies in human genetics, with many examples gleaned from the author's experiences as a practicing genetic counselor. (ISBN 0072325305) Also included is the Answer Key. (ISBN 0072439009)

## ACKNOWLEDGMENTS

I would like to thank many people for their encouragement and assistance in the production of this Seventh Edition. I especially thank Brian Loehr, my Developmental Editor, for continuous support, enthusiasm, and help in improving the usability of the text. It was also a pleasure to work with many other dedicated and creative people at McGraw-Hill during the production of this book, especially James M. Smith, Thomas Timp, Gloria Schiesl, David Hash, Sandy Ludovissy, Carrie Burger, and Jodi Banowetz. I wish to thank Dr. Michael Gaines of the University of Miami for many comments that helped me improve the textbook and Marion Muskiewicz, Reference Librarian at the University of Massachusetts Lowell, who was an enormous help in my efforts to use the university's electronic library. Many reviewers greatly helped improve the quality of this edition. I specifically wish to thank the following:

### Reviewers of the Seventh Edition

**John Belote**
Syracuse University

**Douglas Coulter**
Saint Louis University

**James M. Freed**
Ohio Wesleyan University

**Elliott S. Goldstein**
Arizona State University

**Keith Hartberg**
Baylor University

**Vincent Henrich**
University of North Carolina at Greensboro

**Mitrick A. Johns**
Northern Illinois University

**Philip Mathis**
Middle Tennessee State University

**Bruce McKee**
University of Tennessee

**Elbert Myles**
Tennessee State University

**John Osterman**
University of Nebraska–Lincoln

**Uwe Pott**
University of Wisconsin–Green Bay

**Ken Spitze**
University of Miami

**Randall G. Terry**
University of Montana

**Michael Wooten**
Auburn University

## Reviewers of the Sixth Edition

**Edward Berger**
Dartmouth

**Deborah C. Clark**
Middle Tennessee State University

**John R. Ellison**
Texas A&M University

**Elliott S. Goldstein**
Arizona State University

**Keith Hartberg**
Baylor University

**David R. Hyde**
University of Notre Dame

**Pauline A. Lizotte**
Northwest Missouri State University

**James J. McGivern**
Gannon University

**Gregory J. Phillips**
Iowa State University

**Mark Sanders**
University of California–Davis

**Ken Spitze**
University of Miami

**Joan M. Stoler**
Massachusetts General Hospital, Harvard Medical School

**Robert J. Wiggers**
Stephen F. Austin State University

**Ronald B. Young**
University of Alabama

Lastly, thanks are due to the many students, particularly those in my Introductory Genetics, Population Biology, Evolutionary Biology, and Graduate Seminar courses, who have helped clarify points, find errors, and discover new and interesting ways of looking at the many topics collectively called genetics.

ROBERT H. TAMARIN
*Lowell, Massachusetts*

# 1

# INTRODUCTION

## STUDY OBJECTIVES

1. To examine a brief overview of the modern history of genetics 3
2. To gain an overview of the topics included in this book—the syllabus of genetics 4
3. To analyze the scientific method 5
4. To look at why certain organisms and techniques have been used preferentially in genetics research 7

## STUDY OUTLINE

**A Brief Overview of the Modern History of Genetics** 3
  Before 1860 3
  1860–1900 3
  1900–1944 3
  1944–Present 4
**The Three General Areas of Genetics** 4
**How Do We Know?** 5
**Why Fruit Flies and Colon Bacteria?** 7
**Techniques of Study** 8
**Classical, Molecular, and Evolutionary Genetics** 9
  Classical Genetics 9
  Molecular Genetics 10
  Evolutionary Genetics 13
*Summary* 14
**Box 1.1 The Lysenko Affair** 6

Chameleon, *Cameleo pardalis.*

(© Art Wolfe/Tony Stone Images.)

Genetics is the study of inheritance in all of its manifestations, from the distribution of human traits in a family pedigree to the biochemistry of the genetic material in our chromosomes—deoxyribonucleic acid, or DNA. It is our purpose in this book to introduce and describe the processes and patterns of inheritance. In this chapter, we present a broad outline of the topics to be covered as well as a summary of some of the more important historical advancements leading to our current understanding of genetics.

## A BRIEF OVERVIEW OF THE MODERN HISTORY OF GENETICS

For a generation of students born at a time when incredible technological advances are commonplace, it is valuable to see how far we have come in understanding the mechanisms of genetic processes by taking a very brief, encapsulated look at the modern history of genetics. Although we could discuss prehistoric concepts of animal and plant breeding and ideas going back to the ancient Greeks, we will restrict our brief look to events beginning with the discovery of cells and microscopes. For our purposes, we divide this recent history into four periods: before 1860, 1860–1900, 1900–1944, and 1944 to the present.

### Before 1860

Before 1860, the most notable discoveries paving the way for our current understanding of genetics were the development of light microscopy, the elucidation of the cell theory, and the publication in 1859 of Charles Darwin's *The Origin of Species.* In 1665, Robert Hooke coined the term *cell* in his studies of cork. Hooke saw, in fact, empty cells observed at a magnification of about thirty power. Between 1674 and 1683, Anton van Leeuwenhoek discovered living organisms (protozoa and bacteria) in rainwater. Leeuwenhoek was a master lens maker and produced magnifications of several hundred power from single lenses (fig. 1.1). More than a hundred years passed before compound microscopes could equal Leeuwenhoek's magnifications. In 1833, Robert Brown (the discoverer of Brownian motion) discovered the nuclei of cells, and between 1835 and 1839, Hugo von Mohl described mitosis in nuclei. This era ended in 1858, when Rudolf Virchow summed up the concept of the cell theory with his Latin aphorism *omnis cellula e cellula:* all cells come from preexisting cells. Thus, by 1858, biologists had an understanding of the continuity of cells and knew of the cell's nucleus.

### 1860–1900

The period from 1860 to 1900 encompasses the publication of Gregor Mendel's work with pea plants in 1866 to the rediscovery of his work in 1900. It includes the discoveries of chromosomes and their behavior—insights that shed new light on Mendel's research.

From 1879 to 1885, with the aid of new staining techniques, W. Flemming described the chromosomes—first noticed by C. von Nägeli in 1842—including the way they split during division, and the separation of sister chromatids and their movement to opposite poles of the dividing cell during mitosis. In 1888, W. Waldeyer first used the term *chromosome.* In 1875, O. Hertwig described the fusion of sperm and egg to form the zygote. In the 1880s, Theodor Boveri, as well as K. Rabl and E. van Breden, hypothesized that chromosomes are individual structures with continuity from one generation to the next despite their "disappearance" between cell divisions. In 1885, August Weismann stated that inheritance is based exclusively in the nucleus. In 1887, he predicted the occurrence of a reductional division, which we now call meiosis. By 1890, O. Hertwig and T. Boveri had described the process of meiosis in detail.

### 1900–1944

From 1900 to 1944, modern genetics flourished with the development of the chromosomal theory, which showed

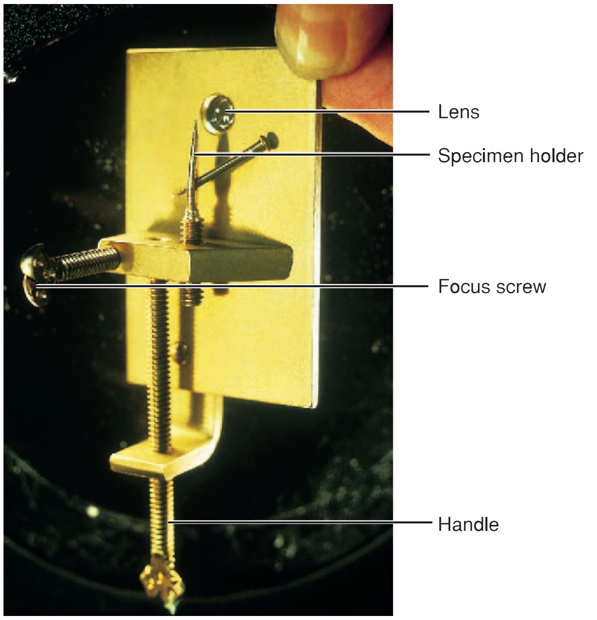


**Figure 1.1** One of Anton van Leeuwenhoek's microscopes, ca. 1680. This single-lensed microscope magnifies up to 200x. (© Kathy Talaro/Visuals Unlimited, Inc.)

that chromosomes are linear arrays of genes. In addition, the foundations of modern evolutionary and molecular genetics were derived.

In 1900, three biologists working independently—Hugo de Vries, Carl Correns, and Erich von Tschermak—rediscovered Mendel's landmark work on the rules of inheritance, published in 1866, thus beginning our era of modern genetics. In 1903, Walter Sutton hypothesized that the behavior of chromosomes during meiosis explained Mendel's rules of inheritance, thus leading to the discovery that genes are located on chromosomes. In 1913, Alfred Sturtevant created the first genetic map, using the fruit fly. He showed that genes existed in a linear order on chromosomes. In 1927, L. Stadler and H. J. Muller showed that genes can be mutated artificially by X rays.

Between 1930 and 1932, R. A. Fisher, S. Wright, and J. B. S. Haldane developed the algebraic foundations for our understanding of the process of evolution. In 1943, S. Luria and M. Delbrück demonstrated that bacteria have normal genetic systems and thus could serve as models for studying genetic processes.

### 1944–Present

The period from 1944 to the present is the era of molecular genetics, beginning with the demonstration that DNA is the genetic material and culminating with our current explosion of knowledge due to recombinant DNA technology.

In 1944, O. Avery and colleagues showed conclusively that deoxyribonucleic acid—DNA—was the genetic material. James Watson and Francis Crick worked out the structure of DNA in 1953. Between 1968 and 1973, W. Arber, H. Smith, and D. Nathans, along with their colleagues, discovered and described restriction endonucleases, the enzymes that opened up our ability to manipulate DNA through recombinant DNA technology. In 1972, Paul Berg was the first to create a recombinant DNA molecule.

Since 1972, geneticists have cloned numerous genes. Scientists now have the capability to create transgenic organisms, organisms with functioning foreign genes. For example, we now have farm animals that produce pharmaceuticals in their milk that are harvested easily and inexpensively for human use. In 1997, the first mammal was cloned, a sheep named Dolly. The sequence of the entire human genome was determined in 2000; we will spend the next century mining its information in the newly created field of genomics, the study of the complete genetic complement of an organism. Although no inherited disease has yet been cured by genetic intervention, we are on the verge of success in numerous diseases, including cancer.

The material here is much too brief to convey any of the detail or excitement surrounding the discoveries of modern genetics. Throughout this book, we will expand on the discoveries made since Darwin first published his book on evolutionary theory in 1859 and since Mendel was rediscovered in 1900.

## THE THREE GENERAL AREAS OF GENETICS

Historically, geneticists have worked in three different areas, each with its own particular problems, terminology, tools, and organisms. These areas are classical genetics, molecular genetics, and evolutionary genetics. In *classical genetics,* we are concerned with the chromosomal theory of inheritance; that is, the concept that genes are

**Table 1.1** **The Three Major Areas of Genetics—Classical, Molecular, and Evolutionary—and the Topics They Cover**

| Classical Genetics | Molecular Genetics | Evolutionary Genetics |
|---|---|---|
| Mendel's principles | Structure of DNA | Quantitative genetics |
| Meiosis and mitosis | Chemistry of DNA | Hardy-Weinberg equilibrium |
| Sex determination | Transcription | Assumptions of equilibrium |
| Sex linkage | Translation | Evolution |
| Chromosomal mapping | DNA cloning and genomics | Speciation |
| Cytogenetics (chromosomal changes) | Control of gene expression | |
| | DNA mutation and repair | |
| | Extrachromosomal inheritance | |

located in a linear fashion on chromosomes and that the relative positions of genes can be determined by their frequency in offspring. *Molecular genetics* is the study of the genetic material: its structure, replication, and expression, as well as the information revolution emanating from the discoveries of recombinant DNA techniques (genetic engineering, including the Human Genome Project). *Evolutionary genetics* is the study of the mechanisms of evolutionary change, or changes in gene frequencies in populations. Darwin's concept of evolution by natural selection finds a firm genetic footing in this area of the study of inheritance (table 1.1).

Today these areas are less clearly defined because of advances made in molecular genetics. Information coming from the study of molecular genetics allows us to understand better the structure and functioning of chromosomes on the one hand and the mechanism of natural selection on the other. In this book we hope to bring together this information from a historical perspective. From Mendel's work in discovering the rules of inheritance (chapter 2) to genetic engineering (chapter 13) to molecular evolution (chapter 21), we hope to present a balanced view of the various topics that make up genetics.

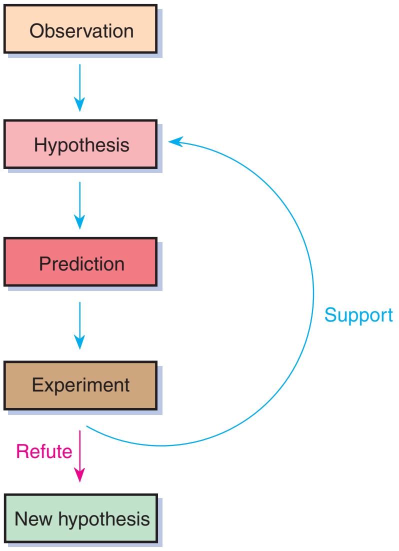


**Figure 1.2** A schematic of the scientific method. An observation leads the researcher to propose a hypothesis, and then to make predictions from the hypothesis and to test these predictions by experiment. The results of the experiment either support or refute the hypothesis. If the experiment refutes the hypothesis, a new hypothesis must be developed. If the experiment supports the hypothesis, the researcher or others design further experiments to try to disprove it.

## HOW DO WE KNOW?

Genetics is an empirical science, which means that our information comes from observations of the natural world. The *scientific method* is a tool for understanding these observations (fig. 1.2). At its heart is the experiment, which tests a guess, called a hypothesis, about how something works. In a good experiment, only two types of outcomes are possible: outcomes that support the hypothesis and outcomes that refute it. Scientists say these outcomes provide *strong inference.*

For example, you might have the idea that organisms can inherit acquired characteristics, an idea put forth by Jean-Baptiste Lamarck (1744–1829), a French biologist. Lamarck used the example of short-necked giraffes evolving into the long-necked giraffes we know of today. He suggested that giraffes that reached higher into trees to get at edible leaves developed longer necks. They passed on these longer necks to their offspring (in small increments in each generation), leading to today's long-necked giraffes. An alternative view, *evolution by natural selection,* was put forward in 1859 by Charles Darwin. According to the Darwinian view, giraffes normally varied in neck length, and these variations were inherited. Giraffes with slightly longer necks would be at an advantage in reaching edible leaves in trees. Therefore, over time, the longer-necked giraffes would survive and reproduce better than the shorter-necked ones. Thus, longer necks would come to predominate. Any genetic *mutations* (changes) that introduced greater neck length would be favored.

To test Lamarck's hypothesis, you might begin by designing an experiment. You could do the experiment on giraffes to test Lamarck's hypothesis directly; however, giraffes are difficult to acquire, maintain, and breed. Remember, though, that you are testing a general hypothesis about the inheritance of acquired characteristics rather than a specific hypothesis about giraffes. Thus, if you are clever enough, you can test the hypothesis with almost any organism. You would certainly choose one that is easy to maintain and manipulate experimentally. Later, you can verify the generality of any particular conclusions with tests on other organisms.

You might decide to use lab mice, which are relatively inexpensive to obtain and keep and have a relatively short generation time of about six weeks, compared with the giraffe's gestation period of over a year. Instead of looking at neck length, you might simply cut off the tip of the tail of each mouse (in a painless manner), using shortened tails as the acquired characteristic. You could then

**BOX 1.1**

## Ethics and Genetics

### *The Lysenko Affair*

As the pictures of geneticists throughout this book indicate, science is a very human activity; people living within societies explore scientific ideas and combine their knowledge. The society in which a scientist lives can affect not only how that scientist perceives the world, but also what that scientist can do in his or her scholarly activities. For example, the United States and other countries decided that mapping the entire human genome would be valuable (see chapter 13). Thus, granting agencies have directed money in this direction. Since much of scientific research is expensive, scientists often can only study areas for which funding is available. Thus, many scientists are working on the Human Genome Project. That is a positive example of society directing research. Examples also exist in which a societal decision has had negative consequences for both the scientific establishment and the society itself. An example is the Lysenko affair in the former Soviet Union during Stalin's and Krushchev's reigns.

Trofim Denisovich Lysenko was a biologist in the former Soviet Union researching the effects of temperature on plant development. At the same time, the preeminent Soviet geneticist was Nikolai Vavilov. Vavilov was interested in improving Soviet crop yields by growing and mating many varieties and selecting the best to be the breeding stock of the next generation. This is the standard way of improving a plant crop or livestock breed (see chapter 18, "Quantitative Inheritance"). The method conforms to genetic principles and therefore is successful. However, it is a slow process that only gradually improves yields.

Lysenko suggested that crop yields could be improved quickly by the inheritance of acquired characteristics (see chapter 21, "Evolution and Speciation"). Although doomed to fail because they denied the true and correct mechanisms of inheritance, Lysenko's ideas were greeted with much enthusiasm by the political elite. The enthusiasm was due not only to the fact that Lysenko promised immediate improvements in crop yields, but also to the fact that Lysenkoism was politically favored. That is, Lysenkoism fit in very well with communism; it promised that nature could be manipulated easily and immediately. If people could manipulate nature so easily, then communism could easily convert people to its doctrines.

Not only did Stalin favor Lysenkoism, but Lysenko himself was favored politically over Vavilov because Lysenko came from peasant stock, whereas Vavilov was from a wealthy family. (Remember that communism

mate these short-tailed mice to see if their offspring have shorter tails. If they do not, you could conclude that a shortened tail, an acquired characteristic, is not inherited. If, however, the next generation of mice have tails shorter than those of their parents, you could conclude that acquired characteristics can be inherited.

One point to note is that every good experiment has a *control,* a part of the experiment that ensures that some unknown variable, often specific to a particular time and place, is not causing the observed changes. For example, in your experiment, the particular food the mice ate may have had an effect on their growth, resulting in offspring with shorter tails. To control for this, you could handle a second group of mice in the exact same way that the experimental mice are handled, except you would not cut off their tails. Any reduction in the lengths of the tails of the offspring of the control mice would indicate an artifact of the experiment rather than the inheritance of acquired characteristics.

The point of doing this experiment (with the control group), as trivial as it might seem, is to determine the answer to a question using data based on what happens in nature. If you design your experiment correctly and carry it out without error, you can be confident about your results. If your results are negative, as ours would be here, then you would reject your hypothesis. Testing hypotheses and rejecting those that are refuted is the essence of the scientific method.

In fact, most of us live our lives according to the scientific method without really thinking about it. For example, we know better than to step out into traffic without looking because we are aware, from experience (observation, experimentation), of the validity of the laws of physics. Although from time to time anti-intellectual movements spread through society, few people actually give up relying on their empirical knowledge of the world to survive (box 1.1).

Nothing in this book is inconsistent with the scientific method. Every fact has been gained by experiment or observation in the real world. If you do not accept something said herein, you can go back to the *original literature,* the published descriptions of original experiwas a revolution of the working class over the wealthy aristocracy.) Supported by Stalin, and then Krushchev, Lysenko gained inordinate power in his country. All visible genetic research in the former Soviet Union was forced to conform to Lysenko's Lamarckian views. People who disagreed with him were forced out of power; Vavilov was arrested in 1940 and died in prison in 1943. It was not until Nikita Krushchev lost power in 1964 that Lysenkoism fell out of favor. Within months, Lysenko's failed pseudoscience was repudiated and Soviet genetics got back on track.

For thirty years, Soviet geneticists were forced into fruitless endeavors, forced out of genetics altogether, or punished for their heterodox views. Superb scientists died in prison while crop improvement programs failed, all because the Soviet dictators favored Lysenkoism. The message of this affair is clear: Politicians can support research that agrees with their political agenda and punish scientists doing research that disagrees with this agenda, but politicians cannot change the truth of the laws of nature. Science, to be effective, must be done in a climate of open inquiry and free expression of ideas. The scientific method cannot be subverted by political bullies.

Trofim Denisovich Lysenko (1898–1976) shows branched wheat to collective farmers in the former Soviet Union. (© SOVFOTO.)

ments in scientific journals (as cited at the end of the book) and read about the work yourself. If you still don't believe a conclusion, you can repeat the work in question either to verify or challenge it. This is in keeping with the nature of the scientific method.

As mentioned, the results of experimental studies are usually published in scientific journals. Examples of journals that many geneticists read include *Genetics, Proceedings of the National Academy of Sciences, Science, Nature, Evolution, Cell, American Journal of Human Genetics, Journal of Molecular Biology,* and hundreds more. The reported research usually undergoes a process called *peer review* in which other scientists review an article before it is published to ensure its accuracy and its relevance. Scientific articles usually include a detailed justification for the work, an outline of the methods that allows other scientists to repeat the work, the results, a discussion of the significance of the results, and citations of prior work relevant to the present study.

At the end of this book, we cite journal articles describing research that has contributed to each chapter. (In chapter 2, we reprint part of Gregor Mendel's work, and in chapter 9, we reprint a research article by J. Watson and F. Crick in its entirety.) We also cite secondary sources, that is, journals and books that publish syntheses of the literature rather than original contributions. These include *Scientific American, Annual Review of Biochemistry, Annual Review of Genetics, American Scientist,* and others. You are encouraged to look at all of these sources in your efforts both to improve your grasp of genetics and to understand how science progresses.

## WHY FRUIT FLIES AND COLON BACTERIA?

As you read this book, you will see that certain organisms are used repeatedly in genetic experiments. If the goal of science is to uncover generalities about the living world, why do geneticists persist in using the same few organisms

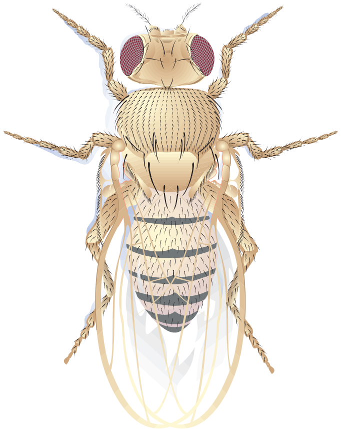


**Figure 1.3** Adult female fruit fly, *Drosophila melanogaster.* Mutations of eye color, bristle type and number, and wing characteristics are easily visible when they occur.

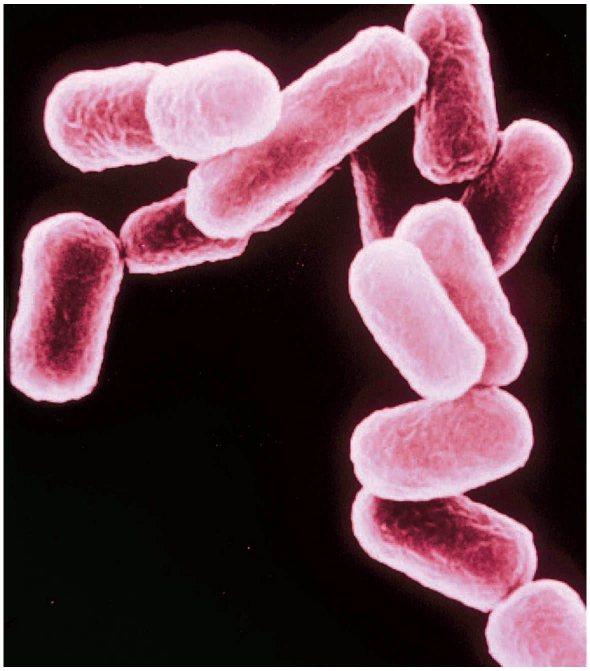


**Figure 1.4** Scanning electron micrograph of *Escherichia coli* bacteria. These rod-shaped bacilli are magnified 18,000x.
(© K. G. Murti/Visuals Unlimited, Inc.)

in their work? The answer is probably obvious: the organisms used for any particular type of study have certain attributes that make them desirable *model organisms* for that research.

In the early stages of genetic research, at the turn of the century, no one had yet developed techniques to do genetic work with microorganisms or mammalian cells. At that time, the organism of preference was the fruit fly, *Drosophila melanogaster,* which developmental biologists had used (fig. 1.3). It has a relatively short generation time of about two weeks, survives and breeds well in the lab, has very large chromosomes in some of its cells, and has many aspects of its *phenotype* (appearance) genetically controlled. For example, it is easy to see the external results of mutations of genes that control eye color, bristle number and type, and wing characteristics such as shape or vein pattern in the fruit fly.

At the middle of this century, when geneticists developed techniques for genetic work on bacteria, the common colon bacterium, *Escherichia coli,* became a favorite organism of genetic researchers (fig. 1.4). Because it had a generation time of only twenty minutes and only a small amount of genetic material, many research groups used it in their experiments. Still later, bacterial viruses, called *bacteriophages,* became very popular in genetics labs. The viruses are constructed of only a few types of protein molecules and a very small amount of genetic material. Some can replicate a hundredfold in an hour. Our point is not to list the major organisms geneticists use, but to suggest why they use some so commonly. Comparative studies are usually done to determine which generalities discovered in the elite genetic organisms are really scientifically universal.

## TECHNIQUES OF STUDY

Each area of genetics has its own particular techniques of study. Often the development of a new technique, or an improvement in a technique, has opened up major new avenues of research. As our technology has improved over the years, geneticists and other scientists have been able to explore at lower and lower levels of biological organization. Gregor Mendel, the father of genetics, did simple breeding studies of plants in a garden at his monastery in Austria in the middle of the nineteenth century. Today, with modern biochemical and biophysical techniques, it has become routine to determine the sequence of *nucleotides* (molecular subunits of DNA and RNA) that make up any particular gene. In fact, one of the most ambitious projects ever carried out in genetics is the mapping of the human genome, all 3.3 billion nucleotides that make up our genes. Only recently was the technology available to complete a project of this magnitude.

## CLASSICAL, MOLECULAR, AND EVOLUTIONARY GENETICS

In the next three sections, we briefly outline the general subject areas covered in the book: classical, molecular, and evolutionary genetics.

### *Classical Genetics*

Gregor Mendel discovered the basic rules of transmission genetics in 1866 by doing carefully controlled breeding experiments with the garden pea plant, *Pisum sativum.* He found that traits, such as pod color, were controlled by genetic elements that we now call *genes* (fig. 1.5). Alternative forms of a gene are called *alleles.* Mendel also discovered that adult organisms have two copies of each gene (*diploid* state); gametes receive just one of these copies (*haploid* state). In other words, one of the two parental copies segregates into any given gamete. Upon fertilization, the zygote gets one copy from each gamete, reconstituting the diploid number (fig. 1.6). When Mendel looked at the inheritance of several

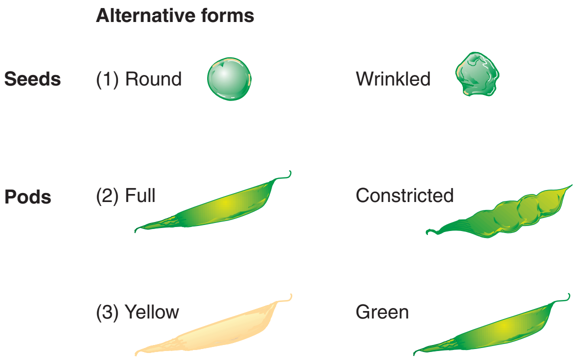

Alternative forms
Seeds
(1) Round
Wrinkled
Pods
(2) Full
Constricted
(3) Yellow
Green

**Figure 1.5** Mendel worked with garden pea plants. He observed seven traits of the plant—each with two discrete forms—that affected attributes of the seed, the pod, and the stem. For example, all plants had either round or wrinkled seeds, full or constricted pods, or yellow or green pods.

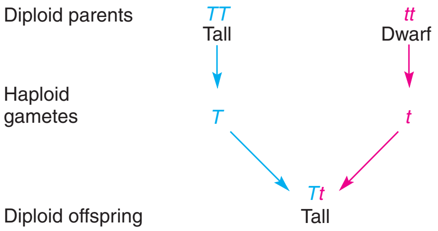

Diploid parents
TT
Tall
tt
Dwarf
Haploid gametes
T
t
Tt
Tall
Diploid offspring

**Figure 1.6** Mendel crossed tall and dwarf pea plants, demonstrating the rule of segregation. A diploid individual with two copies of the gene for tallness (*T*) per cell forms gametes that all have the *T* allele. Similarly, an individual that has two copies of the gene for shortness (*t*) forms gametes that all have the *t* allele. Cross-fertilization produces zygotes that have both the *T* and *t* alleles. When both forms are present (*Tt*), the plant is tall, indicating that the *T* allele is *dominant* to the *recessive* *t* allele.

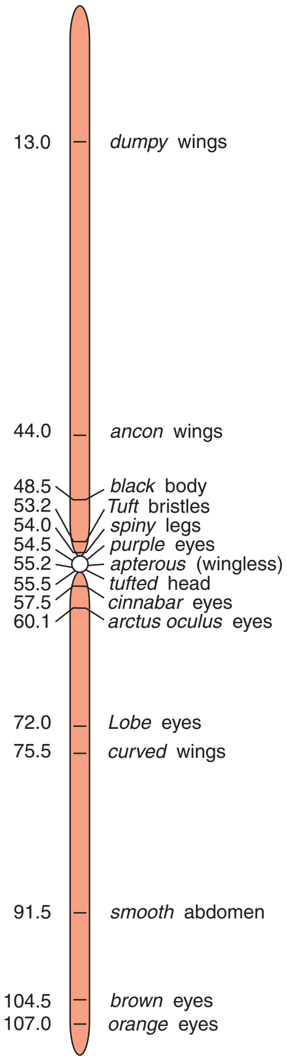

13.0 dumpy wings
44.0 ancon wings
48.5 black body
53.2 Tuft bristles
54.0 spiny legs
54.5 purple eyes
55.2 apterous (wingless)
55.5 tufted head
57.5 cinnabar eyes
60.1 arctus oculus eyes
72.0 Lobe eyes
75.5 curved wings
91.5 smooth abdomen
104.5 brown eyes
107.0 orange eyes

**Figure 1.7** Genes are located in linear order on chromosomes, as seen in this diagram of chromosome 2 of *Drosophila melanogaster,* the common fruit fly. The centromere is a constriction in the chromosome. The numbers are map units.

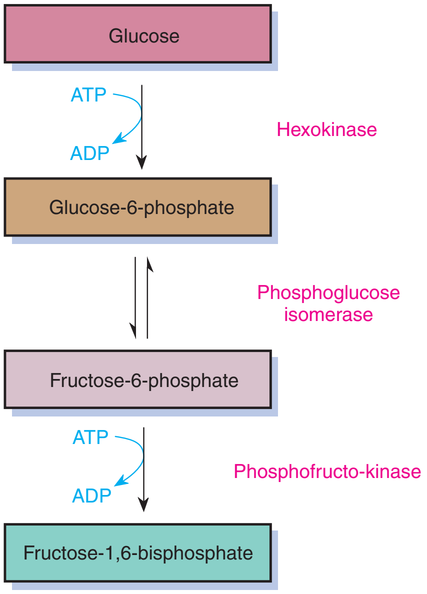


**Figure 1.8** Biochemical pathways are the sequential changes that occur in compounds as cellular reactions modify them. In this case, we show the first few steps in the glycolytic pathway that converts glucose to energy. The pathway begins when glucose + ATP is converted to glucose-6-phosphate + ADP with the aid of the enzyme hexokinase. The enzymes are the products of genes.

traits at the same time, he found that they were inherited independently of each other. His work has been distilled into two rules, referred to as *segregation* and *independent assortment.* Scientists did not accept Mendel's work until they developed an understanding of the segregation of chromosomes during the latter half of the nineteenth century. At that time, in the year 1900, the science of genetics was born.

During much of the early part of this century, geneticists discovered many genes by looking for changed organisms, called *mutants.* Crosses were made to determine the genetic control of mutant traits. From this research evolved chromosomal mapping, the ability to locate the relative positions of genes on chromosomes by crossing certain organisms. The proportion of recombinant offspring, those with new combinations of parental alleles, gives a measure of the physical separation between genes on the same chromosomes in distances called *map units.* From this work arose the chromosomal theory of inheritance: Genes are located at fixed positions on chromosomes in a linear order (fig. 1.7, p. 9). This "beads on a string" model of gene arrangement was not modified to any great extent until the middle of this century, after Watson and Crick worked out the structure of DNA.

In general, genes function by controlling the synthesis of proteins called *enzymes* that act as biological catalysts in biochemical pathways (fig. 1.8). G. Beadle and E. Tatum suggested that one gene controls the formation of one enzyme. Although we now know that many proteins are made up of subunits—the products of several genes—and that some genes code for proteins that are not enzymes and other genes do not code for proteins, the *one-gene-one-enzyme* rule of thumb serves as a general guideline to gene action.

## *Molecular Genetics*

With the exception of some viruses, the genetic material of all cellular organisms is double-stranded DNA, a double helical molecule shaped like a twisted ladder. The backbones of the helices are repeating units of sugars (deoxyribose) and phosphate groups. The rungs of the

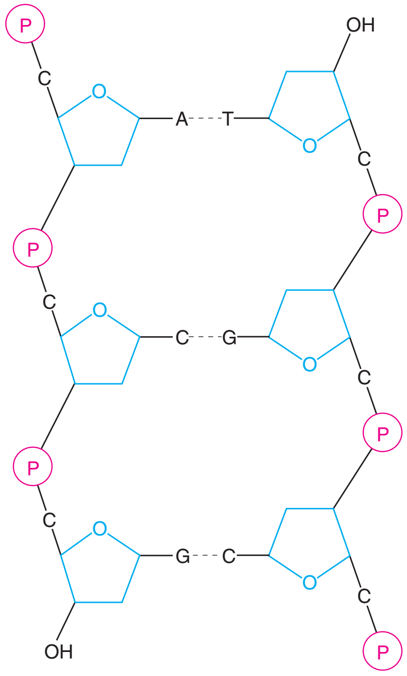


**Figure 1.9** A look at a DNA double helix, showing the sugar-phosphate units that form the molecule's "backbone" and the base pairs that make up the "rungs." We abbreviate a phosphate group as a "P" within a circle; the pentagonal ring containing an oxygen atom is the sugar deoxyribose. Bases are either adenine, thymine, cytosine, or guanine (A, T, C, G).

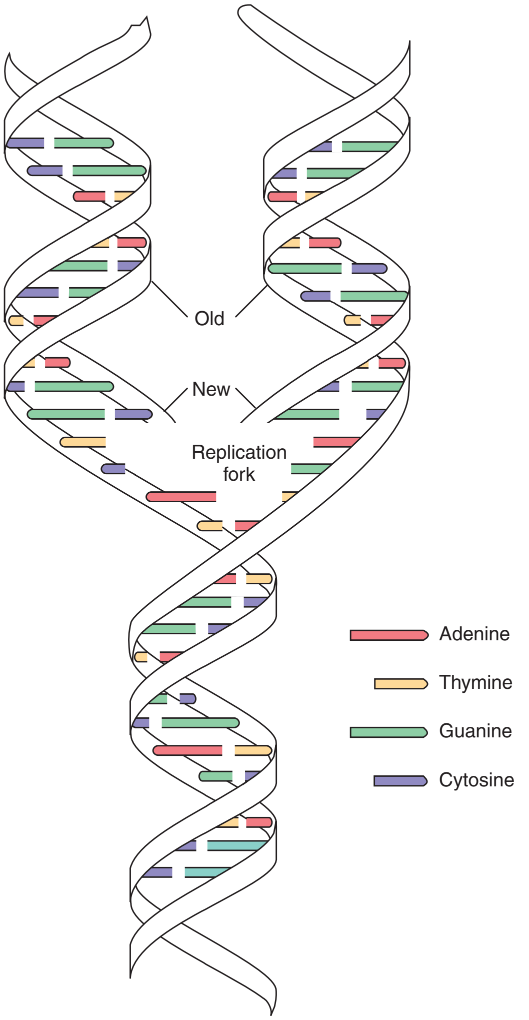


**Figure 1.10** The DNA double helix unwinds during replication, and each half then acts as a template for a new double helix. Because of the rules of complementarity, each new double helix is identical to the original, and the two new double helices are identical to each other. Thus, an AT base pair in the original DNA double helix replicates into two AT base pairs, one in each of the daughter double helices.

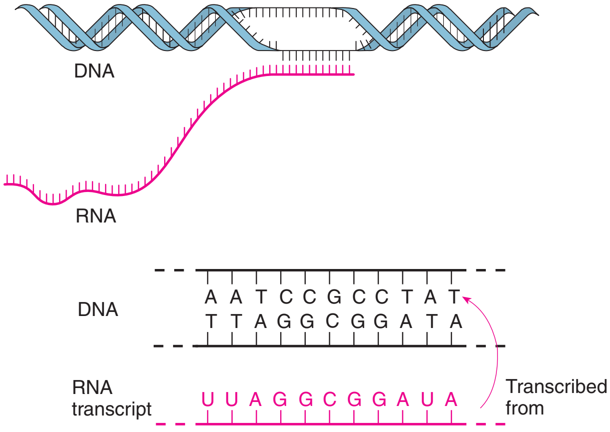


**Figure 1.11** Transcription is the process that synthesizes RNA from a DNA template. Synthesis proceeds with the aid of the enzyme RNA polymerase. The DNA double helix partially unwinds during this process, allowing the base sequence of one strand to serve as a template for RNA synthesis. Synthesis follows the rules of DNA-RNA complementarity: A, T, G, and C of DNA pair with U, A, C, and G, respectively, in RNA. The resulting RNA base sequence is identical to the sequence that would form if the DNA were replicating instead, with the exception that RNA replaces thymine (T) with uracil (U).

ladder are base pairs, with one base extending from each backbone (fig. 1.9). Only four bases normally occur in DNA: adenine, thymine, guanine, and cytosine, abbreviated A, T, G, and C, respectively. There is no restriction on the order of bases on one strand. However, a relationship called *complementarity* exists between bases forming a rung. If one base of the pair is adenine, the other must be thymine; if one base is guanine, the other must be cytosine. James Watson and Francis Crick deduced this structure in 1953, ushering in the era of molecular genetics.

The complementary nature of the base pairs of DNA made the mode of replication obvious to Watson and Crick: The double helix would "unzip," and each strand would act as a template for a new strand, resulting in two double helices exactly like the first (fig. 1.10). Mutation, a change in one of the bases, could result from either an error in base pairing during replication or some damage to the DNA that was not repaired by the time of the next replication cycle.

Information is encoded in DNA in the sequence of bases on one strand of the double helix. During gene expression, that information is *transcribed* into RNA, the other form of nucleic acid, which actually takes part in protein synthesis. RNA differs from DNA in several respects: it has the sugar ribose in place of deoxyribose; it has the base uracil (U) in place of thymine (T); and it usually occurs in a single-stranded form. RNA is transcribed from DNA by the enzyme *RNA polymerase,* using DNA-RNA rules of complementarity: A, T, G, and C in DNA pair with U, A, C, and G, respectively, in RNA (fig. 1.11). The DNA information that is transcribed into RNA codes for the amino acid sequence of proteins. Three nucleotide bases form a *codon* that specifies one of the twenty

**Table 1.2** **The Genetic Code Dictionary of RNA**

| Codon | Amino Acid | Codon | Amino Acid | Codon | Amino Acid | Codon | Amino Acid |
|---|---|---|---|---|---|---|---|
| UUU | Phe | UCU | Ser | UAU | Tyr | UGU | Cys |
| UUC | Phe | UCC | Ser | UAC | Tyr | UGC | Cys |
| UUA | Leu | UCA | Ser | UAA | STOP | UGA | STOP |
| UUG | Leu | UCG | Ser | UAG | STOP | UGG | Trp |
| CUU | Leu | CCU | Pro | CAU | His | CGU | Arg |
| CUC | Leu | CCC | Pro | CAC | His | CGC | Arg |
| CUA | Leu | CCA | Pro | CAA | Gln | CGA | Arg |
| CUG | Leu | CCG | Pro | CAG | Gln | CGG | Arg |
| AUU | Ile | ACU | Thr | AAU | Asn | AGU | Ser |
| AUC | Ile | ACC | Thr | AAC | Asn | AGC | Ser |
| AUA | Ile | ACA | Thr | AAA | Lys | AGA | Arg |
| AUG | Met (START) | ACG | Thr | AAG | Lys | AGG | Arg |
| GUU | Val | GCU | Ala | GAU | Asp | GGU | Gly |
| GUC | Val | GCC | Ala | GAC | Asp | GGC | Gly |
| GUA | Val | GCA | Ala | GAA | Glu | GGA | Gly |
| GUG | Val | GCG | Ala | GAG | Glu | GGG | Gly |

*Note:* A codon, specifying one amino acid, is three bases long (read in RNA bases in which U replaced the T of DNA). There are sixty-four different codons, specifying twenty naturally occurring amino acids (abbreviated by three letters: e.g., Phe is phenylalanine—see fig. 11.1 for the names and structures of the amino acids). Also present is *stop* (UAA, UAG, UGA) and *start* (AUG) information.

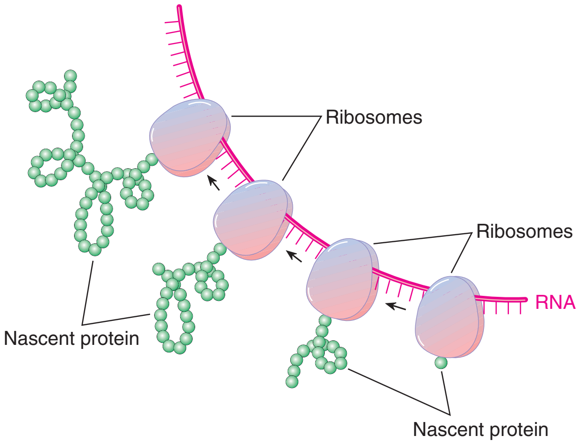


**Figure 1.12** In prokaryotes, RNA translation begins shortly after RNA synthesis. A ribosome attaches to the RNA and begins reading the RNA codons. As the ribosome moves along the RNA, amino acids attach to the growing protein. When the process is finished, the completed protein is released from the ribosome, and the ribosome detaches from the RNA. As the first ribosome moves along, a second ribosome can attach at the beginning of the RNA, and so on, so that an RNA strand may have many ribosomes attached at one time.

naturally occurring amino acids used in protein synthesis. The sequence of bases making up the codons are referred to as the genetic code (table 1.2).

The process of *translation,* the decoding of nucleotide sequences into amino acid sequences, takes place at the ribosome, a structure found in all cells that is made up of RNA and proteins (fig. 1.12). As the RNA moves along the ribosome one codon at a time, one amino acid attaches to the growing protein for each codon.

The major control mechanisms of gene expression usually act at the transcriptional level. For transcription to take place, the RNA polymerase enzyme must be able to pass along the DNA; if this movement is prevented, transcription stops. Various proteins can bind to the DNA, thus preventing the RNA polymerase from continuing, providing a mechanism to control transcription. One particular mechanism, known as the *operon model,* provides the basis for a wide range of control mechanisms in prokaryotes and viruses. Eukaryotes generally contain no operons; although we know quite a bit about some control systems for eukaryotic gene expression, the general rules are not as simple.

In recent years, there has been an explosion of information resulting from *recombinant DNA techniques.* This revolution began with the discovery of *restriction endonucleases,* enzymes that cut DNA at specific sequences. Many of these enzymes leave single-stranded ends on the cut DNA. If a restriction enzyme acts on both a *plasmid,* a small, circular extrachromosomal unit found in some bacteria, and another piece of DNA (called foreign DNA), the two will be left with identical single-stranded free ends. If the cut plasmid and cut foreign DNA are mixed together, the free ends can re-form double helices, and the plasmid can take in a single piece of foreign DNA (fig. 1.13). Final repair processes create a completely closed circle of DNA. The hybrid plasmid is then reinserted into the bacterium. When the bacterium grows, it replicates the plasmid DNA, producing many copies of the foreign DNA. From that point, the foreign DNA can be isolated and sequenced, allowing researchers to determine the exact order of bases making up the foreign DNA. (In 2000, scientists announced the complete sequencing of the human genome.) That sequence can tell us much about how a gene works. In addition, the foreign genes can function within the bacterium, resulting in bacteria expressing the foreign genes and producing their protein products. Thus we have, for example, *E. coli* bacteria that produce human growth hormone.

This technology has tremendous implications in medicine, agriculture, and industry. It has provided the opportunity to locate and study disease-causing genes, such as the genes for cystic fibrosis and muscular dystrophy, as well as suggesting potential treatments. Crop plants and farm animals are being modified for better productivity by improving growth and disease resistance. Industries that apply the concepts of genetic engineering are flourishing.

One area of great interest to geneticists is cancer research. We have discovered that a single gene that has lost its normal control mechanisms (an *oncogene*) can cause changes that lead to cancer. These oncogenes exist normally in noncancerous cells, where they are called proto-oncogenes, and are also carried by viruses, where they are called viral oncogenes. Cancer-causing viruses are especially interesting because most of them are of the RNA type. AIDS is caused by one of these RNA viruses, which attacks one of the cells in the immune system. Cancer can also occur when genes that normally prevent cancer, genes called anti-oncogenes, lose function. Discovering the mechanism by which our immune system can produce millions of different protective proteins (*antibodies*) has been another success of modern molecular genetics.

## Evolutionary Genetics

From a genetic standpoint, evolution is the change in allelic frequencies in a population over time. Charles Darwin described evolution as the result of natural selection. In the 1920s and 1930s, geneticists, primarily Sewall Wright, R. A. Fisher, and J. B. S. Haldane, provided algebraic models to describe evolutionary processes. The marriage of Darwinian theory and population genetics has been termed *neo-Darwinism.*

In 1908, G. H. Hardy and W. Weinberg discovered that a simple genetic equilibrium occurs in a population if the population is large, has random mating, and has negligible effects of mutation, migration, and natural selection. This equilibrium gives population geneticists a baseline for comparing populations to see if any evolutionary

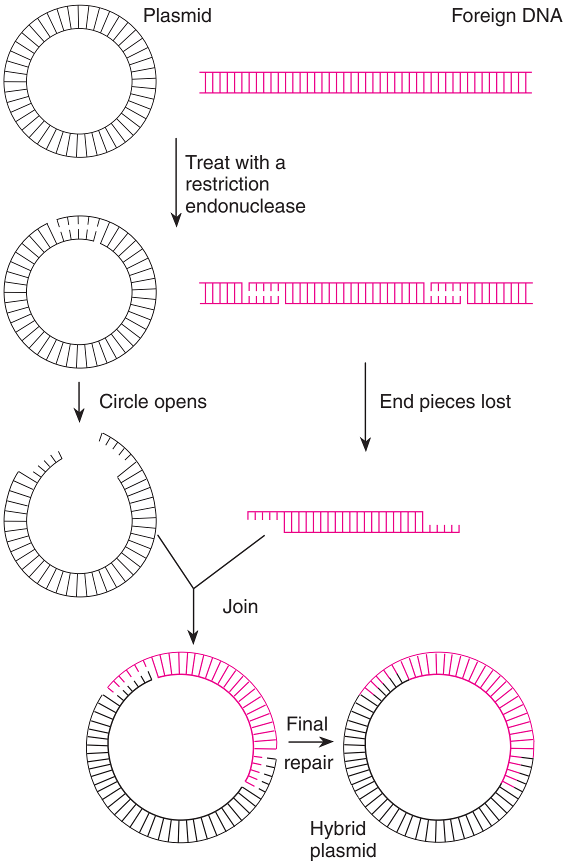


**Figure 1.13** Hybrid DNA molecules can be constructed from a plasmid and a piece of foreign DNA. The ends are made compatible by cutting both DNAs with the same restriction endonuclease, leaving complementary ends. These ends will re-form double helices to form intact hybrid plasmids when the two types of DNA mix. A repair enzyme, DNA ligase, finishes patching the hybrid DNA within the plasmid. The hybrid plasmid is then reinjected into a bacterium, to be grown into billions of copies that will later be available for isolation and sequencing, or the hybrid plasmid can express the foreign DNA from within the host bacterium.

processes are occurring. We can formulate a statement to describe the equilibrium condition: If the assumptions are met, the population will not experience changes in allelic frequencies, and these allelic frequencies will accurately predict the frequencies of *genotypes* (allelic combinations in individuals, e.g., *AA, Aa,* or *aa*) in the population.

Recently, several areas of evolutionary genetics have become controversial. Electrophoresis (a method for separating proteins and other molecules) and subsequent DNA sequencing have revealed that much more *polymorphism* (variation) exists within natural populations than older mathematical models could account for. One of the more interesting explanations for this variability is that it is neutral. That is, natural selection, the guiding force of evolution, does not act differentially on many, if not most, of the genetic differences found so commonly in nature. At first, this theory was quite controversial, attracting few followers. Now it seems to be the view the majority accept to explain the abundance of molecular variation found in natural populations.

Another controversial theory concerns the rate of evolutionary change. It is suggested that most evolutionary change is not gradual, as the fossil record seems to indicate, but occurs in short, rapid bursts, followed by long periods of very little change. This theory is called *punctuated equilibrium.*

A final area of evolutionary biology that has generated much controversy is the theory of *sociobiology.* Sociobiologists suggest that social behavior is under genetic control and is acted upon by natural selection, as is any morphological or physiological trait. This idea is controversial mainly as it applies to human beings; it calls altruism into question and suggests that to some extent we are genetically programmed to act in certain ways. People have criticized the theory because they feel it justifies racism and sexism.

## SUMMARY

The purpose of this chapter has been to provide a brief history of genetics and a brief overview of the following twenty chapters. We hope it serves to introduce the material and to provide a basis for early synthesis of some of the material that, of necessity, is presented in the discrete units called chapters. This chapter also differs from all the others because it lacks some of the end materials that should be of value to you as you proceed: solved problems, and exercises and problems. These features are presented chapter by chapter throughout the remainder of the book. At the end of the book, we provide answers to exercises and problems and a glossary of all **boldface** words throughout the book.

*Suggested Readings for chapter 1 are on page B-1.*

# 2

# MENDEL'S PRINCIPLES

## STUDY OBJECTIVES

1. To understand that genes are discrete units that control the appearance of an organism 17
2. To understand Mendel's rules of inheritance: segregation and independent assortment 18
3. To understand that dominance is a function of the interaction of alleles; similarly, epistasis is a function of the interaction of nonallelic genes 22
4. To define how genes generally control the production of enzymes and thus the fate of biochemical pathways 37

## STUDY OUTLINE

**Mendel's Experiments** 17
**Segregation** 18
  Rule of Segregation 18
  Testing the Rule of Segregation 21
**Dominance Is Not Universal** 22
**Nomenclature** 23
**Multiple Alleles** 25
**Independent Assortment** 26
  Rule of Independent Assortment 27
  Testcrossing Multihybrids 30
**Genotypic Interactions** 30
  Epistasis 32
  Mechanism of Epistasis 34
**Biochemical Genetics** 37
  Inborn Errors of Metabolism 37
  One-Gene-One-Enzyme Hypothesis 38
*Summary* 40
*Solved Problems* 40
*Exercises and Problems* 41
*Critical Thinking Questions* 45
**Box 2.1 Excerpts from Mendel's Original Paper** 28
**Box 2.2 Did Mendel Cheat?** 30

The garden pea plant, *Pisum sativum.*

(© Adam Hart-Davis/SPL/Photo Researchers, Inc.)

Genetics is concerned with the transmission, expression, and evolution of genes, the molecules that control the function, development, and ultimate appearance of individuals. In this section of the book, we will look at the rules of transmission that govern genes and affect their passage from one generation to the next. Gregor Johann Mendel discovered these rules of inheritance; we derive and expand upon his rules in this chapter (fig. 2.1).

In 1900, three botanists, Carl Correns of Germany, Erich von Tschermak of Austria, and Hugo de Vries of Holland, defined the rules governing the transmission of traits from parent to offspring. Some historical controversy exists as to whether these botanists actually rediscovered Mendel's rules by their own research or whether their research led them to Mendel's original paper. In any case, all three made important contributions to the early stages of genetics. The rules had been published previously, in 1866, by an obscure Austrian monk, Gregor Johann Mendel. Although his work was widely available after 1866, the scientific community was not ready to appreciate Mendel's great contribution until the turn of the century. There are at least four reasons for this lapse of thirty-four years.

**Figure 2.1** Gregor Johann Mendel (1822–84). (Reproduced by permission of the Moravski Museum, Mendelianum.)

First, before Mendel's experiments, biologists were primarily concerned with explaining the transmission of characteristics that could be measured on a continuous scale, such as height, cranium size, and longevity. They were looking for rules of inheritance that would explain such **continuous variations,** especially after Darwin put forth his theory of evolution in 1859 (see chapter 21). Mendel, however, suggested that inherited characteristics were discrete and constant **(discontinuous):** peas, for example, were either yellow or green. Thus, evolutionists were looking for small changes in traits with continuous variation, whereas Mendel presented them with rules for discontinuous variation. His principles did not seem to apply to the type of variation that biologists thought prevailed. Second, there was no physical element identified with Mendel's inherited entities. One could not say, upon reading Mendel's work, that a certain subunit of the cell followed Mendel's rules. Third, Mendel worked with large numbers of offspring and converted these numbers to ratios. Biologists, practitioners of a very descriptive science at the time, were not well trained in mathematical tools. And last, Mendel was not well known and did not persevere in his attempts to convince the academic community that his findings were important.

Between 1866 and 1900, two major changes took place in biological science. First, by the turn of the century, not only had scientists discovered chromosomes, but they also had learned to understand chromosomal movement during cell division. Second, biologists were better prepared to handle mathematics by the turn of the century than they were during Mendel's time.

## MENDEL'S EXPERIMENTS

Gregor Mendel was an Austrian monk (of Brünn, Austria, which is now Brno, Czech Republic). In his experiments, he tried to **crossbreed** plants that had discrete, nonoverlapping characteristics and then to observe the distribution of these characteristics over the next several generations. Mendel worked with the common garden pea plant, *Pisum sativum.* He chose the pea plant for at least three reasons: (1) The garden pea was easy to cultivate and had a relatively short life cycle. (2) The plant had discontinuous characteristics such as flower color and pea texture. (3) In part because of its anatomy, pollination of the plant was easy to control. Foreign pollen could be kept out, and **cross-fertilization** could be accomplished artificially.

Figure 2.2 shows a cross section of the pea flower that indicates the keel, in which the male and female parts develop. Normally, **self-fertilization** occurs when pollen falls onto the stigma before the bud opens. Mendel cross-fertilized the plants by opening the keel of

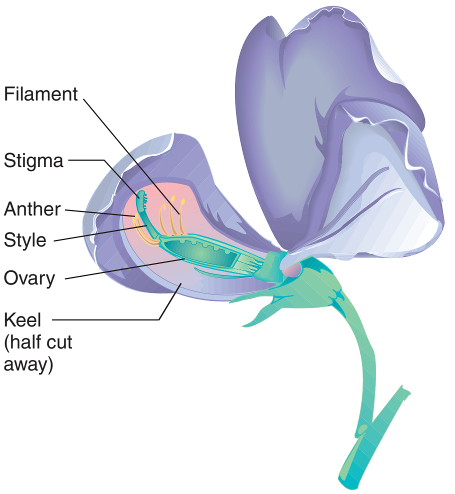


**Figure 2.2** Anatomy of the garden pea plant flower. The female part, the pistil, is composed of the stigma, its supporting style, and the ovary. The male part, the stamen, is composed of the pollen-producing anther and its supporting filament.

a flower before the anthers matured and placing pollen from another plant on the stigma. In the more than ten thousand plants Mendel examined, only a few were fertilized other than the way he had intended (either self- or cross-pollinated).

Mendel used plants obtained from suppliers and grew them for two years to ascertain that they were homogeneous, or true-breeding, for the particular characteristic under study. He chose for study the seven characteristics shown in figure 2.3. Take as an example the characteristic of plant height. Although height is often continuously distributed, Mendel used plants that displayed only two alternatives: tall or dwarf. He made the crosses shown in figure 2.4. In the parental, or $P_1$, generation, dwarf plants pollinated tall plants, and, in a **reciprocal cross,** tall plants pollinated dwarf plants, to determine whether the results were independent of the parents' sex. As we will see later on, some traits follow inheritance patterns related to the sex of the parent carrying the traits. In those cases, reciprocal crosses give different results; with Mendel's tall and dwarf pea plants, the results were the same.

Offspring of the cross of $P_1$ individuals are referred to as the first **filial generation,** or $F_1$. Mendel also referred to them as **hybrids** because they were the offspring of unlike parents (tall and dwarf). We will specifically refer to the offspring of tall and dwarf peas as **monohybrids** because they are hybrid for only one characteristic (height). Since all the $F_1$ offspring plants were tall, Mendel referred to tallness as the **dominant** trait. The alternative, dwarfness, he referred to as **recessive.** Different forms of a gene that exist within a population are termed **alleles.** The terms *dominant* and *recessive* are used to describe both the relationship between the alleles and the traits they control. Thus, we say that both the allele for tallness and the trait, tall, are dominant. Dominance applies to the appearance of the trait when both a dominant and a recessive allele are present. It does not imply that the dominant trait is better, is more abundant, or will increase over time in a population.

When the $F_1$ offspring of figure 2.4 were self-fertilized to produce the $F_2$ generation, both tall and dwarf offspring occurred; the dwarf characteristic reappeared. Among the $F_2$ offspring, Mendel observed 787 tall and 277 dwarf plants for a ratio of 2.84:1. It is an indication of Mendel's insight that he recognized in these numbers an approximation to a 3:1 ratio, a ratio that suggested to him the mechanism of inheritance at work in pea plant height.

## SEGREGATION

### *Rule of Segregation*

Mendel assumed that each plant contained two determinants (which we now call **genes**) for the characteristic of height. For example, a hybrid $F_1$ pea plant possesses the dominant allele for tallness and the recessive allele for dwarfness for the gene that determines plant height. A pair of alleles for dwarfness is required to develop the recessive phenotype. Only one of these alleles is passed into a single **gamete,** and the union of two gametes to form a zygote restores the double complement of alleles. The fact that the recessive trait reappears in the $F_2$ generation shows that the allele controlling it was hidden in the $F_1$ individual and passed on unaffected. This explanation of the passage of discrete trait determinants, or genes, comprises Mendel's first principle, the **rule of segregation.** The rule of segregation can be summarized as follows: A gamete receives only one allele from the pair of alleles an organism possesses; fertilization (the union of two gametes) reestablishes the double number. We can visualize this process by redrawing figure 2.4 using letters to denote the alleles. Mendel used capital letters to denote alleles that control dominant traits and lowercase letters for alleles that control recessive traits. Following this notation, $T$ refers to the allele controlling tallness and $t$ refers to the allele controlling shortness (dwarf stature). From figure 2.5, we can see that Mendel's rule of segregation explains the homogeneity of the $F_1$ generation (all tall) and the 3:1 ratio of tall-to-dwarf offspring in the $F_2$ generation.

Let us define some terms. The **genotype** of an organism is the gene combination it possesses. In figure 2.5,

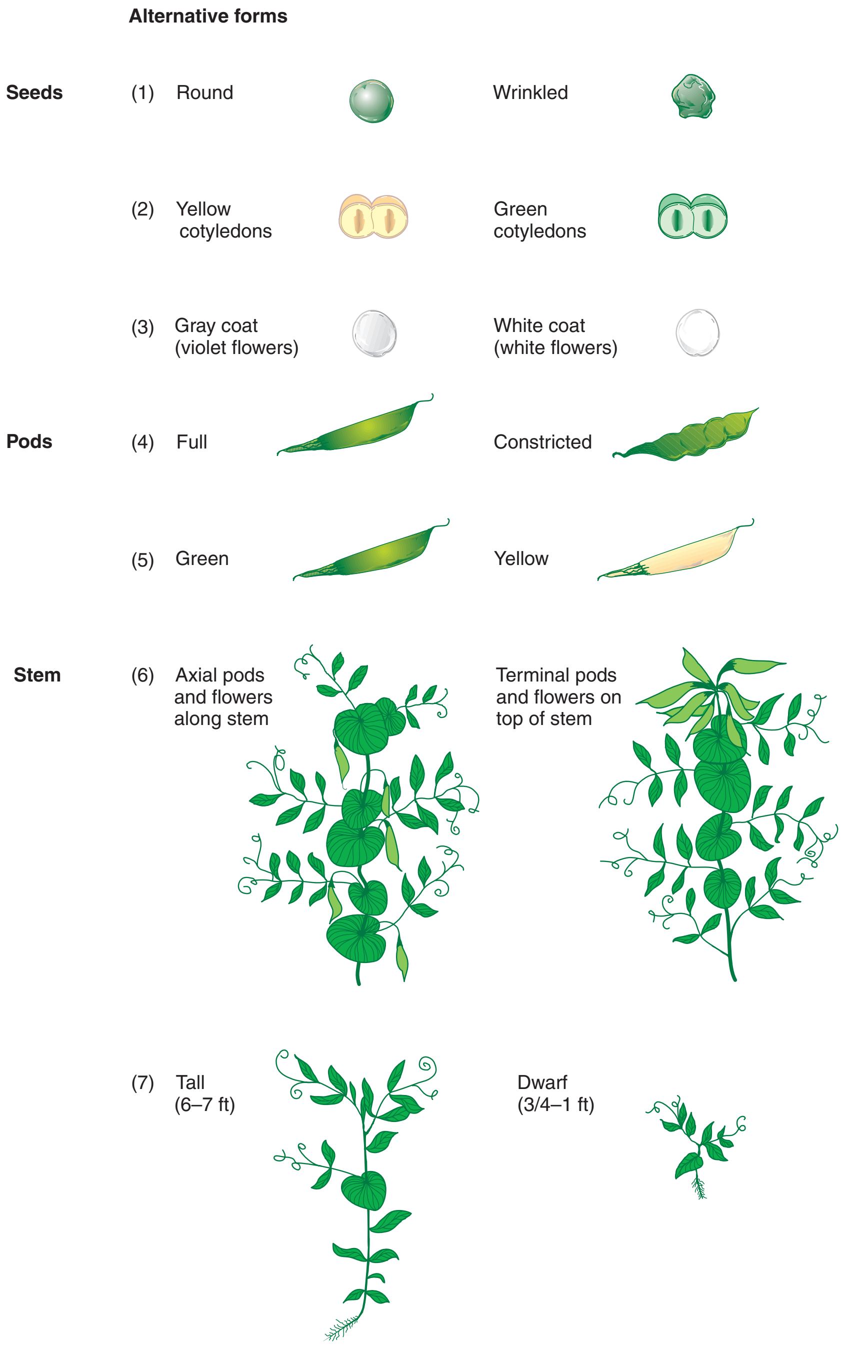

Alternative forms
Seeds
(1) Round
Wrinkled
(2) Yellow cotyledons
Green cotyledons
(3) Gray coat (violet flowers)
White coat (white flowers)
Pods
(4) Full
Constricted
(5) Green
Yellow
Stem
(6) Axial pods and flowers along stem
Terminal pods and flowers on top of stem
(7) Tall (6–7 ft)
Dwarf (3/4–1 ft)

**Figure 2.3** Seven characteristics that Mendel observed in peas. Traits in the left column are dominant.

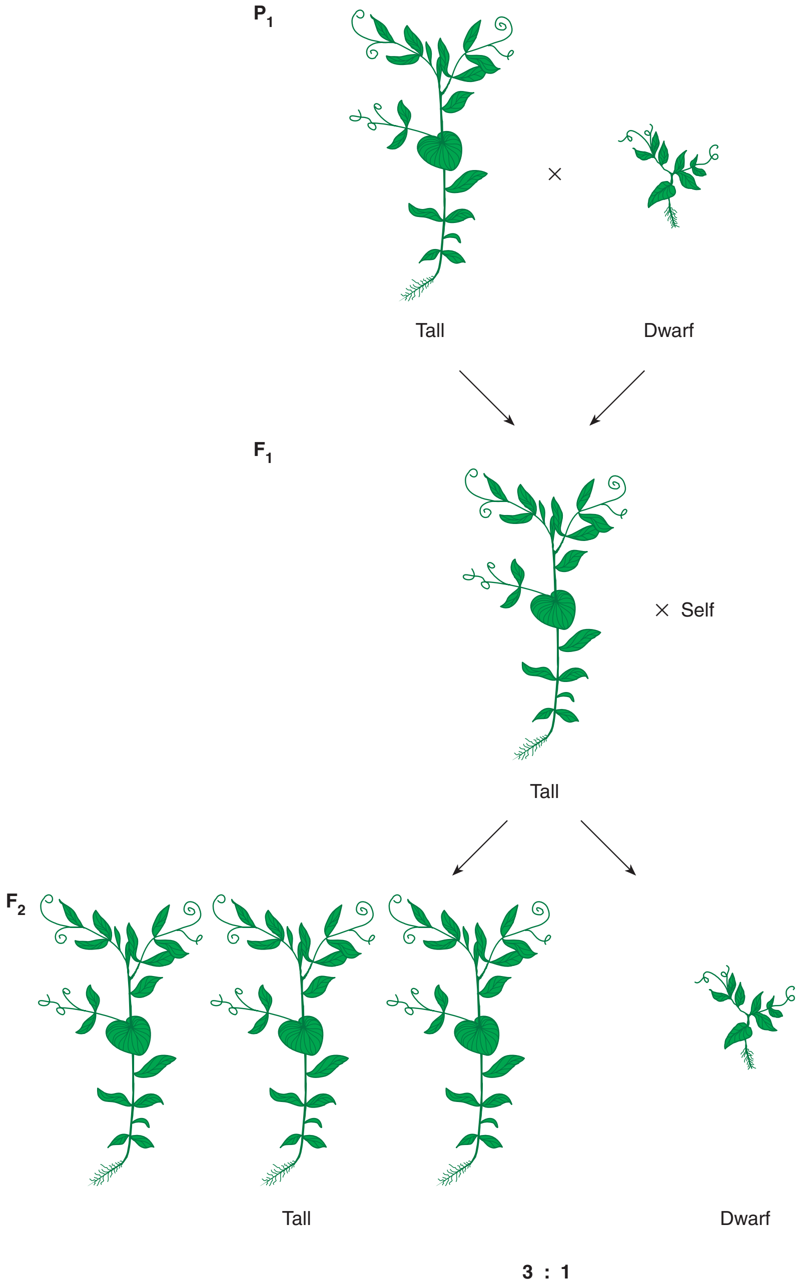

$P_1$
×
Tall
Dwarf
$F_1$
× Self
Tall
$F_2$
Tall
Dwarf
3 : 1

**Figure 2.4** First two offspring generations from the cross of tall plants with dwarf plants.

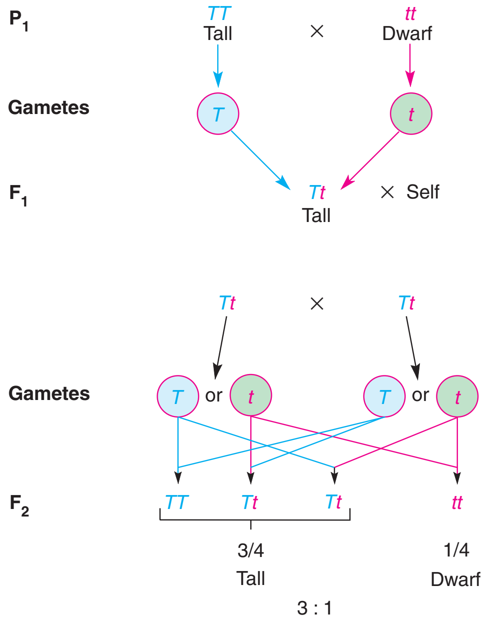
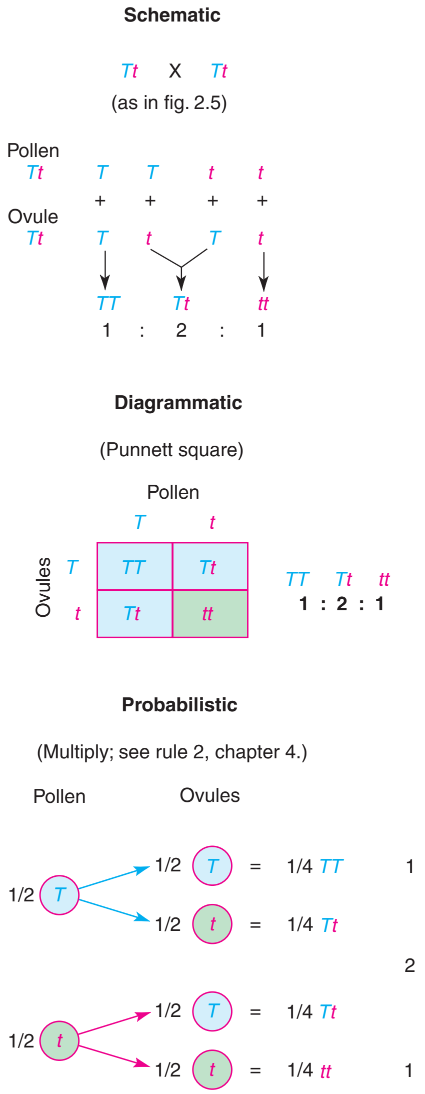

P1
TT
Tall
×
tt
Dwarf
Gametes
T
t
F1
Tt
Tall
× Self
Tt
×
Tt
Gametes
T
or
t
T
or
t
F2
TT
Tt
Tt
tt
3/4
Tall
1/4
Dwarf
3 : 1

**Figure 2.5** Assigning genotypes to the cross in figure 2.4.


Schematic
Tt X Tt
(as in fig. 2.5)
Pollen
Tt
T T t t
+ + + +
Ovule
Tt
T t T t
TT Tt tt
1 : 2 : 1
Diagrammatic
(Punnett square)
Pollen
T t
Ovules
T
t
TT Tt tt
1 : 2 : 1
Probabilistic
(Multiply; see rule 2, chapter 4.)
Pollen
Ovules
1/2 T
1/2 T = 1/4 TT
1/2 t = 1/4 Tt
1/2 t
1/2 T = 1/4 Tt
1/2 t = 1/4 tt
1
2
1

| | T | t |
|---|---|---|
| T | TT | Tt |
| t | Tt | tt |

**Figure 2.6** Methods of determining $F_2$ genotypic combinations in a self-fertilized monohybrid. The Punnett square diagram is named after the geneticist Reginald C. Punnett.

the genotype of the parental tall plant is *TT;* that of the $F_1$ tall plant is *Tt.* **Phenotype** refers to the observable attributes of an organism. Plants with either of the two genotypes *TT* or *Tt* are phenotypically tall. Genotypes come in two general classes: **homozygotes,** in which both alleles are the same, as in *TT* or *tt,* and **heterozygotes,** in which the two alleles are different, as in *Tt.* William Bateson coined these last two terms in 1901. Danish botanist Wilhelm Johannsen first used the word *gene* in 1909.

If we look at figure 2.5, we can see that the *TT* homozygote can produce only one type of gamete, the *T*-bearing kind, and the *tt* homozygote can similarly produce only *t*-bearing gametes. Thus, the $F_1$ individuals are uniformly heterozygous *Tt,* and each $F_1$ individual can produce two kinds of gametes in equal frequencies, *T*- or *t*-bearing. In the $F_2$ generation, these two types of gametes randomly pair during fertilization. Figure 2.6 shows three ways of picturing this process.

## *Testing the Rule of Segregation*

We can see from figure 2.6 that the $F_2$ generation has a phenotypic ratio of 3:1, the classic Mendelian ratio. However, we also see a genotypic ratio of 1:2:1 for dominant homozygote:heterozygote:recessive homozygote. Demonstrating this genotypic ratio provides a good test of Mendel's rule of segregation.

The simplest way to test the hypothesis is by **progeny testing,** that is, by self-fertilizing $F_2$ individuals to produce an $F_3$ generation, which Mendel did (fig. 2.7). Treating the rule of segregation as a hypothesis, it is possible to predict the frequencies of the phenotypic classes that would result. The dwarf $F_2$ plants should be recessive homozygotes, and so, when **selfed** (self-fertilized), they should produce only *t*-bearing gametes and only dwarf offspring in the $F_3$ generation. The tall $F_2$ plants, however, should be a heterogeneous group, one-third of which should be homozygous *TT* and two-thirds heterozygous *Tt.* The tall homozygotes, when selfed, should produce only tall $F_3$ offspring (genotypically *TT*). However, the $F_2$ heterozygotes, when selfed, should produce

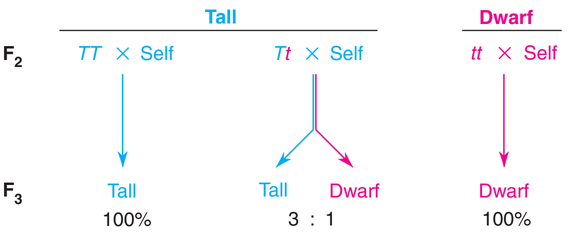


**Figure 2.7** Mendel self-fertilized $F_2$ tall and dwarf plants. He found that all the dwarf plants produced only dwarf progeny. Among the tall plants, 72% produced both tall and dwarf progeny in a 3:1 ratio.

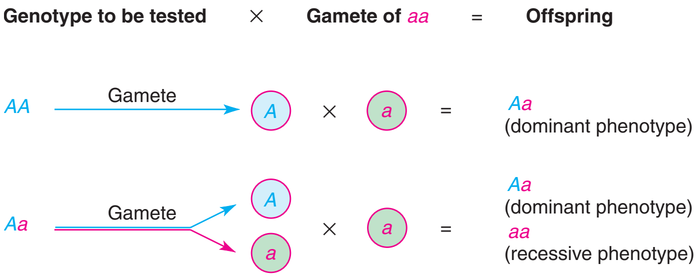


**Figure 2.8** Testcross. In a testcross, the phenotype of an offspring is determined by the allele the offspring inherits from the parent with the genotype being tested.

tall and dwarf offspring in a ratio identical to that the selfed $F_1$ plants produced: three tall to one dwarf offspring. Mendel found that all the dwarf (homozygous) $F_2$ plants bred true as predicted. Among the tall, 28% (28/100) bred true (produced only tall offspring) and 72% (72/100) produced both tall and dwarf offspring. Since the prediction was one-third (33.3%) and two-thirds (66.7%), respectively, Mendel's observed values were very close to those predicted. We thus conclude that Mendel's progeny-testing experiment confirmed his hypothesis of segregation. In fact, a statistical test—developed in chapter 4—would also the support this conclusion.

Another way to test the segregation rule is to use the extremely useful method of the **testcross,** that is, a cross of any organism with a recessive homozygote. (Another type of cross, a **backcross,** is the cross of a progeny with an individual that has a parental genotype. Hence, a testcross can often be a backcross.) Since the gametes of the recessive homozygote contain only recessive alleles, the alleles that the gametes of the other parent carry will determine the phenotypes of the offspring. If a gamete from the organism being tested contains a recessive allele, the resulting $F_1$ organism will have a recessive phenotype; if it contains a dominant allele, the $F_1$ organism will have a dominant phenotype. Thus, in a testcross, the genotypes of the gametes from the organism being tested determine the phenotypes of the offspring (fig. 2.8). A testcross of the tall $F_2$ plants in figure 2.5 would produce the results shown in figure 2.9. These results further confirm Mendel's rule of segregation.

**Tall (two classes)**

$TT \times tt$ = all $Tt$
$Tt \times tt = Tt : tt$
**1 : 1**

**Figure 2.9** Testcrossing the dominant phenotype of the $F_2$ generation from figure 2.5.

## DOMINANCE IS NOT UNIVERSAL

If dominance were universal, the heterozygote would always have the same phenotype as the dominant homozygote, and we would always see the 3:1 ratio when heterozygotes are crossed. If, however, the heterozygote were distinctly different from both homozygotes, we

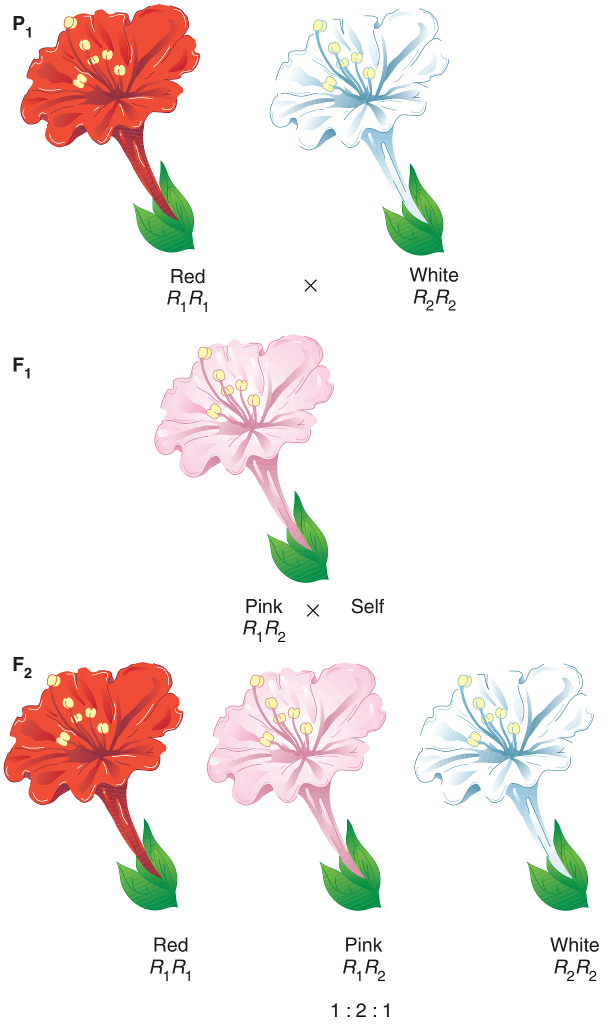


**Figure 2.10** Flower color inheritance in the four-o'clock plant: an example of partial, or incomplete, dominance.

would see a 1:2:1 ratio of phenotypes when heterozygotes are crossed. In **partial dominance** (or **incomplete dominance**), the phenotype of the heterozygote falls between those of the two homozygotes. An example occurs in flower petal color in some plants.

Using four-o'clock plants (*Mirabilis jalapa*), we can cross a plant that has red flower petals with another that has white flower petals; the offspring will have pink flower petals. If these pink-flowered $F_1$ plants are crossed, the $F_2$ plants appear in a ratio of 1:2:1, having red, pink, or white flower petals, respectively (fig. 2.10). The pink-flowered plants are heterozygotes that have a petal color intermediate between the red and white colors of the homozygotes. In this case, one allele ($R_1$) specifies red pigment color, and another allele specifies no color ($R_2$; the flower petals have a white background color). Flowers in heterozygotes ($R_1R_2$) have about half the red pigment of the flowers in red homozygotes ($R_1R_1$) because the heterozygotes have only one copy of the allele that produces color, whereas the homozygotes have two copies.

As technology has improved, we have found more and more cases in which we can differentiate the heterozygote. It is now clear that dominance and recessiveness are phenomena dependent on which alleles are interacting and on what phenotypic level we are studying. For example, in Tay-Sachs disease, homozygous recessive children usually die before the age of three after suffering severe nervous system degeneration; heterozygotes seem to be normal. As biologists have discovered how the disease works, they have made the detection of the heterozygotes possible.

As with many genetic diseases, the culprit is a defective **enzyme** (protein catalyst). Afflicted homozygotes have no enzyme activity, heterozygotes have about half the normal level, and, of course, homozygous normal individuals have the full level. In the case of Tay-Sachs disease, the defective enzyme is hexoseaminidase-A, needed for proper lipid metabolism. Modern techniques allow technicians to assay the blood for this enzyme and to identify heterozygotes by their intermediate level of enzyme activity. Two heterozygotes can now know that there is a 25% chance that any child they bear will have the disease. They can make an educated decision as to whether or not to have children.

The other category in which the heterozygote is discernible occurs when the heterozygous phenotype is not on a scale somewhere between the two homozygotes, but actually expresses both phenotypes simultaneously. We refer to this situation as **codominance.** For example, people with blood type AB are heterozygotes who express both the *A* and *B* alleles for blood type (see the section entitled "Multiple Alleles" for more information about blood types). Electrophoresis (a technique described in chapter 5) lets us see proteins directly and also gives us many examples of codominance when we can see the protein products of both alleles.

## NOMENCLATURE

Throughout the last century, botanists, zoologists, and microbiologists have adopted different methods for naming alleles. Botanists and mammalian geneticists tend to prefer the capital-lowercase scheme. *Drosophila* geneticists and microbiologists have adopted schemes that relate to the **wild-type.** The wild-type is the phenotype of the organism commonly found in nature. Though other naturally occurring phenotypes of the same species may also be present, there is usually an agreed-upon common

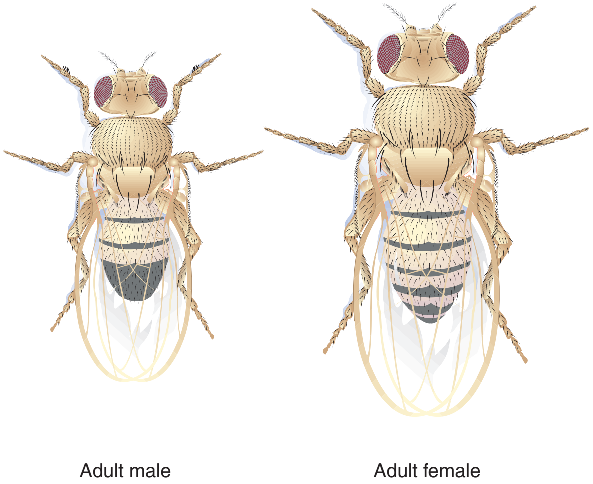


**Figure 2.11** Wild-type fruit fly, *Drosophila melanogaster.*

**Table 2.1** Some Mutants of *Drosophila*

| Mutant Designation | Description | Dominance Relationship to Wild-Type |
|---|---|---|
| abrupt (*ab*) | Shortened, longitudinal, median wing vein | Recessive |
| amber (*amb*) | Pale yellow body | Recessive |
| black (*b*) | Black body | Recessive |
| Bar (*B*) | Narrow, vertical eye | Dominant |
| dumpy (*dp*) | Reduced wings | Recessive |
| Hairless (*H*) | Various bristles absent | Dominant |
| white (*w*) | White eye | Recessive |
| white-apricot ($w^a$) | Apricot-colored eye (allele of white eye) | Recessive |

phenotype that is referred to as the wild-type. For fruit flies (*Drosophila*), organisms commonly used in genetic studies, the wild-type has red eyes and round wings (fig. 2.11). Alternatives to the wild-type are referred to as **mutants** (fig. 2.12). Thus, red eyes are wild-type, and white eyes are mutant. Fruit fly genes are named after the mutant, beginning with a capital letter if the mutation is dominant and a lowercase letter if it is recessive. Table 2.1 gives some examples. The wild-type allele often carries the symbol of the mutant with a + added as a superscript; by definition, every mutant has a wild-type allele as an alternative. For example, *w* stands for the white-eye allele, a recessive mutation. The wild-type (red eyes) is thus assigned the symbol $w^+$. Hairless is a dominant allele with the symbol *H.* Its wild-type allele is denoted as $H^+$. Sometimes geneticists use the + symbol alone for the wild-type, but only when there will be no confusion about its use. If we are discussing eye color only, then + is clearly the same as $w^+$: both mean red eyes. However, if we are discussing both eye color and bristle morphology, the + alone could refer to either of the two aspects of the phenotype and should be avoided.

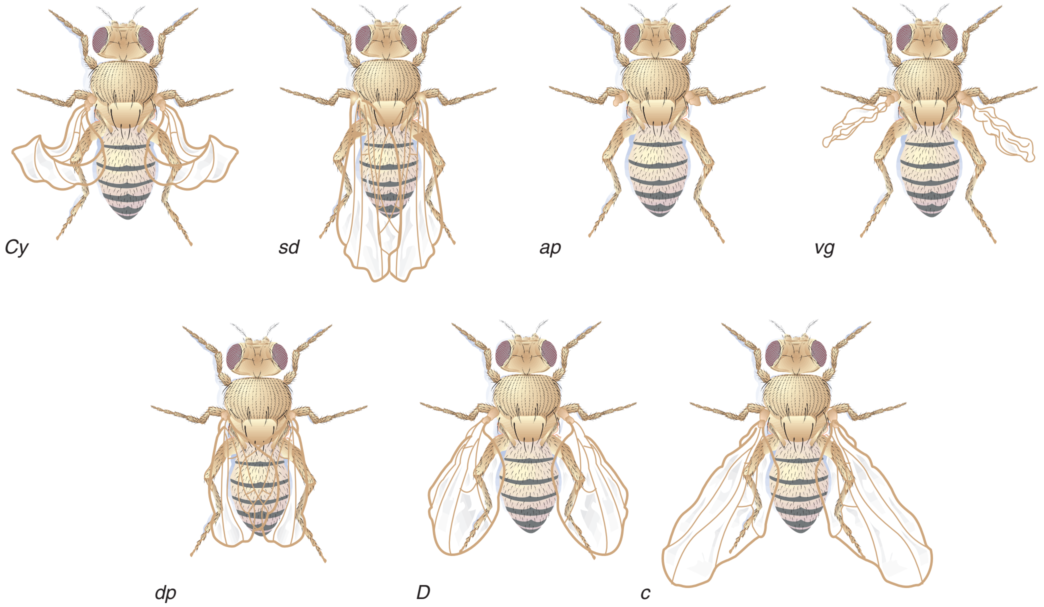


**Figure 2.12** Wing mutants of *Drosophila melanogaster* and their allelic designations: *Cy,* curly; *sd,* scalloped; *ap,* apterous; *vg,* vestigial; *dp,* dumpy; *D,* Dichaete; *c,* curved.

**Table 2.2** **ABO Blood Types with Immunity Reactions**

| Blood Type Corresponding to Antigens on Red Blood Cells | Antibodies in Serum | Genotype | Reaction of Red Cells to Anti-A Antibodies | Reaction of Red Cells to Anti-B Antibodies |
|---|---|---|---|---|
| O | Anti-A and anti-B | $ii$ | − | − |
| A | Anti-B | $I^AI^A$ or $I^Ai$ | + | − |
| B | Anti-A | $I^BI^B$ or $I^Bi$ | − | + |
| AB | None | $I^AI^B$ | + | + |

## MULTIPLE ALLELES

A given gene can have more than two alleles. Although any particular individual can have only two, many alleles of a given gene may exist in a population. The classic example of multiple human alleles is in the ABO blood group, which Karl Landsteiner discovered in 1900. This is the best known of all the red-cell antigen systems primarily because of its importance in blood transfusions. There are four blood-type phenotypes produced by three alleles (table 2.2). The $I^A$ and $I^B$ alleles are responsible for the production of the A and B antigens found on the surface of the erythrocytes (red blood cells). Antigens are substances, normally foreign to the body, that induce the immune system to produce antibodies (proteins that bind to the antigens). The ABO system is unusual because antibodies can be present (e.g., anti-B antibodies can exist in a type A person) without prior exposure to the antigen. Thus, people with a particular ABO antigen on their red cells will have in their serum the antibody against the other antigen: type A persons have A antigen on their red cells and anti-B antibody in their serum; type B persons have B antigen on their red cells and anti-A antibody in their serum; type O persons do not have either antigen but have both antibodies in their serum; and type AB persons have both A and B antigens and form neither anti-A or anti-B antibodies in their serum.

The $I^A$ and $I^B$ alleles, coding for glycosyl transferase enzymes, each cause a different modification to the terminal sugars of a mucopolysaccharide (H structure) found on the surface of red blood cells (fig. 2.13). They are codominant because both modifications (antigens) are present in a heterozygote. In fact, whichever enzyme (product of the $I^A$ or $I^B$ allele) reaches the H structure first will modify it. Once modified, the H structure will not respond to the other enzyme. Therefore, both A and B antigens will be produced in the heterozygote in roughly equal proportions. The $i$ allele causes no change to the H structure: because of a mutation it produces a nonfunctioning enzyme. The $i$ allele and its phenotype are recessive; the presence of the $I^A$ or $I^B$ allele, or both,

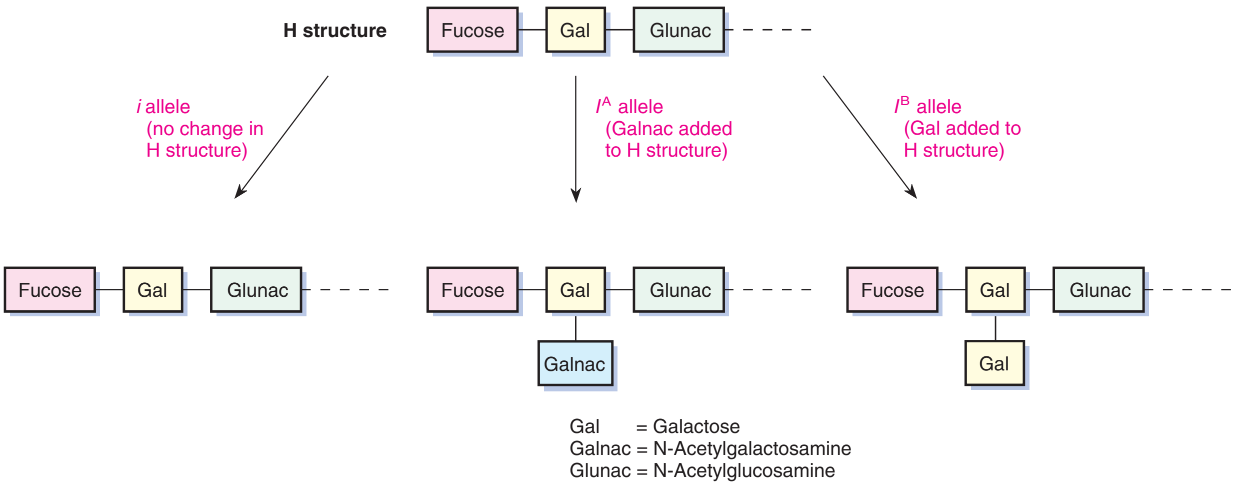

H structure
Fucose
Gal
Glunac
i allele (no change in H structure)
I^A allele (Galnac added to H structure)
I^B allele (Gal added to H structure)
Fucose
Gal
Glunac
Fucose
Gal
Glunac
Galnac
Fucose
Gal
Glunac
Gal
Gal = Galactose
Galnac = N-Acetylgalactosamine
Glunac = N-Acetylglucosamine

**Figure 2.13** Function of the $I^A$, $I^B$, and $i$ alleles of the ABO gene. The gene products of the $I^A$ and $I^B$ alleles of the ABO gene affect the terminal sugars of a mucopolysaccharide (H structure) found on red blood cells. The gene products of the $I^A$ and $I^B$ alleles are the enzymes alpha-3-N-acetyl-D-galactosaminyltransferase and alpha-3-D-galactosyltransferase, respectively.

will modify the H product, thus masking the fact that the *i* allele was ever there.

Adverse reactions to blood transfusions primarily occur because the antibodies in the recipient's serum react with the antigens on the donor's red blood cells. Thus, type A persons cannot donate blood to type B persons. Type B persons have anti-A antibody, which reacts with the A antigen on the donor red cells and causes the cells to clump.

Since both $I^A$ and $I^B$ are dominant to the *i* allele, this system not only shows multiple allelism, it also demonstrates both codominance and simple dominance. (As with virtually any system, intense study yields more information, and subgroups of type A are known. We will not, however, deal with that complexity here.) According to the American Red Cross, 46% of blood donors in the United States are type O, 40% are type A, 10% are type B, and 4% are type AB.

Many other genes also have multiple alleles. In some plants, such as red clover, there is a gene, the S gene, with several hundred alleles that prevent self-fertilization. This means that a pollen grain is not capable of forming a successful pollen tube in the style if the pollen grain or its parent plant has a self-incompatibility allele that is also present in the plant to be fertilized. Thus, pollen grains from a flower falling on its own stigma are rejected. Only a pollen grain with either a different self-incompatibility allele or from a parent plant with different self-incompatibility alleles is capable of fertilization; this avoids inbreeding. Thus, over evolutionary time, there has been selection for many alleles of this gene. Presumably, a foreign plant would not want to be mistaken for the same plant, providing the selective pressure for many alleles to survive in a population. Recent research has indicated that the products of the *S* alleles are ribonuclease enzymes, enzymes that destroy RNA. Researchers are interested in discovering the molecular mechanisms for this pollen rejection.

In *Drosophila,* numerous alleles of the white-eye gene exist, and people have numerous hemoglobin alleles. In fact, multiple alleles are the rule rather than the exception.

## INDEPENDENT ASSORTMENT

Mendel also analyzed the inheritance pattern of traits observed two at a time. He looked, for instance, at plants that differed in the form and color of their peas: he crossed true-breeding (homozygous) plants that had seeds that were round and yellow with plants that produced seeds that were wrinkled and green. Mendel's results appear in figure 2.14. The $F_1$ plants all had round, yellow seeds, which demonstrated that round was dominant to wrinkled and yellow was dominant to green. When these $F_1$ plants were self-fertilized, they produced an $F_2$ generation that had all four possible combinations of the two seed characteristics: round, yellow seeds; round, green seeds; wrinkled, yellow seeds; and wrinkled, green seeds. The numbers Mendel reported in these categories were 315, 108, 101, and 32, respectively. Dividing each number by 32 gives a 9.84 to 3.38 to 3.16 to 1.00 ratio, which is very close to a 9:3:3:1 ratio. As you will see, this is the ratio we would expect if the genes governing these two traits behaved independently of each other.

In figure 2.14, the letter *R* is assigned to the dominant allele, round, and *r* to the recessive allele, wrinkled; *Y* and *y* are used for yellow and green color, respectively. In figure 2.15, we have rediagrammed the cross in figure 2.14. The $P_1$ plants in this cross produce only one type of gamete each, *RY* for the parent with the dominant traits and *ry* for the parent with the recessive traits. The resulting $F_1$ plants are heterozygous for both genes **(dihybrid).** Self-fertilizing the dihybrid (*RrYy*) produces the $F_2$ generation.

In constructing the **Punnett square** in figure 2.15 to diagram the $F_2$ generation, we make a critical assumption: The four types of gametes from each parent will be produced in equal numbers, and hence every offspring category, or "box," in the square is equally likely. Thus, because sixteen boxes make up the Punnett square (named after its inventor, Reginald C. Punnett), the ratio of $F_2$ offspring should be in sixteenths. Grouping the $F_2$ offspring by phenotype, we find there are 9/16 that have round, yellow seeds; 3/16 that have round, green seeds; 3/16 that have wrinkled, yellow seeds; and 1/16 that have wrinkled, green seeds. This is the origin of the expected 9:3:3:1 $F_2$ ratio.

Reginald C. Punnett (1875–1967).
From *Genetics,* 58 (1968): frontispiece.
Courtesy of the Genetics Society of America.

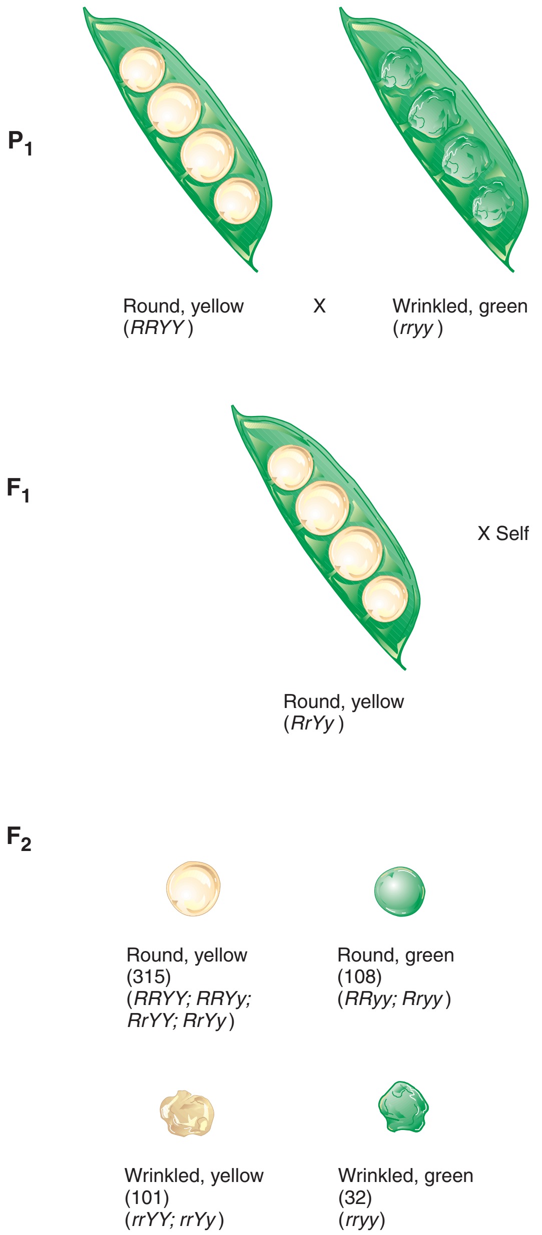

P1
Round, yellow (RRYY)
X
Wrinkled, green (rryy)
F1
X Self
Round, yellow (RrYy)
F2
Round, yellow (315) (RRYY; RRYy; RrYY; RrYy)
Round, green (108) (RRyy; Rryy)
Wrinkled, yellow (101) (rrYY; rrYy)
Wrinkled, green (32) (rryy)

**Figure 2.14** Independent assortment in garden peas.

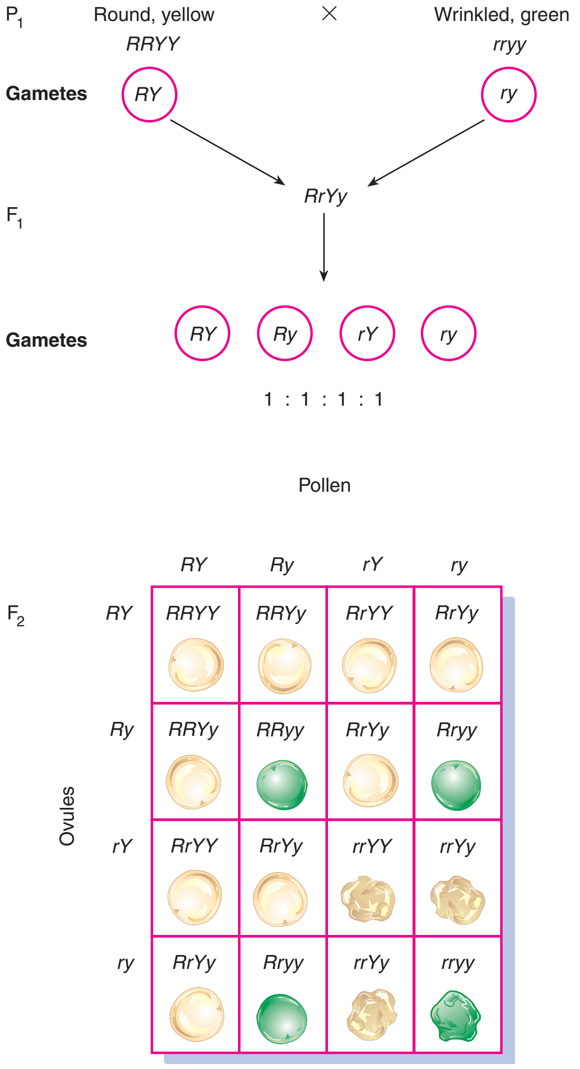

P1
Round, yellow
RRYY
×
Wrinkled, green
rryy
Gametes
RY
ry
F1
RrYy
Gametes
RY
Ry
rY
ry
1 : 1 : 1 : 1
Pollen
RY
Ry
rY
ry
F2
Ovules
RY
RRYY
RRYy
RrYY
RrYy
Ry
RRYy
RRyy
RrYy
Rryy
rY
RrYY
RrYy
rrYY
rrYy
ry
RrYy
Rryy
rrYy
rryy

**Figure 2.15** Assigning genotypes to the cross in figure 2.14.

## *Rule of Independent Assortment*

This ratio comes about because the two characteristics behave independently. The $F_1$ plants produce four types of gametes (check fig. 2.15): *RY, Ry, rY,* and *ry.* These gametes occur in equal frequencies. Regardless of which seed shape allele a gamete ends up with, it has a 50:50 chance of getting either of the alleles for color—the two genes are segregating, or assorting, independently. This is the essence of Mendel's second rule, the **rule of independent assortment,** which states that alleles for one gene can segregate independently of alleles for other genes. Are the alleles for the two characteristics of color and form segregating properly according to Mendel's first principle?

If we look only at seed shape (see fig. 2.14), we find that a homozygote with round seeds was crossed with a homozygote with wrinkled seeds in the $P_1$ generation ($RR \times rr$). This cross yields only heterozygous plants with round seeds (*Rr*) in the $F_1$ generation. When these

## BOX 2.1

### Historical Perspectives

### *Excerpts from Mendel's Original Paper*

In February and March of 1865, Mendel delivered two lectures to the Natural History Society of Brünn. These were published as a single forty-eight-page article handwritten in German. The article appeared in the 1865 *Proceedings of the Society,* which came out in 1866. It was entitled "*Versuche über Pflanzen-Hybriden,*" which means "Experiments in Plant Hybridization." Following are some paragraphs from the English translation to give us some sense of the original.

In his introductory remarks, Mendel writes:

> That, so far, no generally applicable law governing the formation and development of hybrids has been successfully formulated can hardly be wondered at by anyone who is acquainted with the extent of the task, and can appreciate the difficulties with which experiments of this class have to contend. A final decision can only be arrived at when we shall have before us the results of detailed experiments made on plants belonging to the most diverse orders.
>
> Those who survey the work done in this department will arrive at the conviction that among all the numerous experiments made, not one has been carried out to such an extent and in such a way as to make it possible to determine the number of different forms under which the offspring of hybrids appear, or to arrange these forms with certainty according to their separate generations, or definitely to ascertain their statistical relations. . . .
>
> The paper now presented records the results of such a detailed experiment. This experiment was practically confined to a small plant group, and is now, after eight years' pursuit, concluded in all essentials. Whether the plan upon which the separate experiments were conducted and carried out was the best suited to attain the desired end is left to the friendly decision of the reader.

After discussing the origin of his seeds and the nature of the experiments, Mendel discusses the $F_1$, or hybrid, generation:

> This is precisely the case with the Pea hybrids. In the case of each of the seven crosses the hybrid-character resembles that of one of the parental forms so closely that the other either escapes observation completely or cannot be detected with certainty. This circumstance is of great importance in the determination and classification of the forms under which the offspring of the hybrids appear. Henceforth in this paper those characters which are transmitted entire, or almost unchanged in the hybridization, and therefore in themselves constitute the characters of the hybrid, are termed the *dominant,* and those which become latent in the process, *recessive.* The expression "recessive" has been chosen because the characters thereby designated withdraw or entirely disappear in the hybrids, but nevertheless reappear unchanged in their progeny, as will be demonstrated later on.

He then writes about the $F_2$ generation:

> In this generation there reappear, together with the dominant characters, also the recessive ones with their peculiarities fully developed, and this occurs in the definitely expressed average proportion of three to one, so that among each four plants of this generation three display the dominant character and one the recessive. This relates without exception to all the characters which were investigated in the experiments. The angular wrinkled form of the seed, the green colour of the albumen, the white colour of the seed-coats and the flowers, the constrictions of the pods, the yellow colour of the unripe pod, of the stalk, of the calyx, and of the leaf venation, the umbel-like form of the inflorescence, and the dwarfed stem, all reappear in the numerical proportion given, without any essential alteration. *Transitional forms were not observed in any experiment. . . .*

---

$F_1$ plants are self-fertilized, the result is 315 + 108 round seeds (*RR* or *Rr*) and 101 + 32 wrinkled seeds (*rr*) in the $F_2$ generation. This is a 423:133 or a 3.18:1.00 phenotypic ratio—very close to the expected 3:1 ratio. So the gene for seed shape is segregating normally. In a similar manner, if we look only at the gene for color, we see that the $F_2$ ratio of yellow to green seeds is 416:140, or 2.97:1.00—again, very close to a 3:1 ratio. Thus, when two genes are segregating normally according to the rule of segregation, their independent behavior demonstrates the rule of independent assortment (box 2.1).

From the Punnett square in figure 2.15, you can see that because of dominance, all phenotypic classes except the homozygous recessive one—wrinkled, green

> Expt. 1. Form of seed.—From 253 hybrids 7,324 seeds were obtained in the second trial year. Among them were 5,474 round or roundish ones and 1,850 angular wrinkled ones. Therefrom the ratio 2.96 to 1 is deduced.
>
> If $A$ be taken as denoting one of the two constant characters, for instance the dominant, $a$ the recessive, and $Aa$ the hybrid form in which both are conjoined, the expression
>
> $$A + 2Aa + a$$
>
> shows the terms in the series for the progeny of the hybrids of two differentiating characters.

Mendel used a notation system different from ours. He designated heterozygotes with both alleles (e.g., $Aa$) but homozygotes with only one allele or the other (e.g., $A$ for our $AA$). Thus, whereas he recorded $A + 2Aa + a$, we would record $AA + 2Aa + aa$. Mendel then went on to discuss the dihybrids. He mentions the genotypic ratio of 1:2:1:2:4:2:1:2:1 and the principle of independent assortment:

> The fertilized seeds appeared round and yellow like those of the seed parents. The plants raised therefrom yielded seeds of four sorts, which frequently presented themselves in one pod. In all, 556 seeds were yielded by 15 plants, and of those there were:
>
> 315 round and yellow,
> 101 wrinkled and yellow,
> 108 round and green,
> 32 wrinkled and green.
>
> Consequently the offspring of the hybrids, if two kinds of differentiating characters are combined therein, are represented by the expression
>
> $$AB + Ab + aB + ab + 2ABb + 2aBb + 2AaB + 2Aab + 4AaBb.$$

(In today's notation, we would write: $AABB + AAbb + aaBB + aabb + 2AABb + 2aaBb + 2AaBB + 2Aabb + 4AaBb$.)

> This expression is indisputably a combination series in which the two expressions for the characters $A$ and $a$, $B$ and $b$ are combined. We arrive at the full number of the classes of the series by the combination of the expressions
>
> $$A + 2Aa + a$$
>
> $$B + 2Bb + b$$

(In today's notation we would write

$$AA + 2Aa + aa$$

$$BB + 2Bb + bb.)$$

> There is therefore no doubt that for the whole of the characters involved in the experiments the principle applies that *the offspring of the hybrids in which several essentially different characters are combined exhibit the terms of a series of combinations, in which the developmental series for each pair of differentiating characters are united.* It is demonstrated at the same time that *the relation of each pair of different characters in hybrid union is independent of the other differences in the two original parental stocks.*

Table 1 is a summary of all the data Mendel presented on monohybrids (the data from only one dihybrid and one trihybrid cross were presented):

**Table 1** Mendel's Data

| | Dominant Phenotype | Recessive Phenotype | Ratio |
|---|---|---|---|
| Seed form | 5,474 | 1,850 | 2.96:1 |
| Cotyledon color | 6,022 | 2,001 | 3.01:1 |
| Seed coat color | 705 | 224 | 3.15:1 |
| Pod form | 882 | 299 | 2.95:1 |
| Pod color | 428 | 152 | 2.82:1 |
| Flower position | 651 | 207 | 3.14:1 |
| Stem length | 787 | 277 | 2.84:1 |
| Total | 14,949 | 5,010 | 2.98:1 |

*Source*: Copyright The Royal Horticultural Society. Taken from the *Journal of the Royal Horticultural Society*, vol. 26. Pg. 1–32. 1901.

seeds—are actually genetically heterogeneous, with phenotypes made up of several genotypes. For example, the dominant phenotypic class, with round, yellow seeds, represents four genotypes: *RRYY, RRYy, RrYY,* and *RrYy.* When we group all the genotypes by phenotype, we obtain the ratio shown in figure 2.16. Thus, with complete dominance, a self-fertilized dihybrid gives a 9:3:3:1 phenotypic ratio in its offspring ($F_2$). A 1:2:1:2:4:2:1:2:1 genotypic ratio also occurs in the $F_2$ generation. If the two genes exhibited incomplete dominance or codominance, the latter would also be the phenotypic ratio. What ratio would be obtained if one gene exhibited dominance and the other did not? An example of this case appears in figure 2.17.

### BOX 2.2

## Historical Perspectives

### Did Mendel Cheat?

Overwhelming evidence gathered during this century has proven the correctness of Mendel's conclusions. However, close scrutiny of Mendel's paper has led some to suggest that (1) Mendel failed to report the inheritance of traits that did not show independent assortment and (2) Mendel fabricated numbers. Both these claims are, on the surface, difficult to ignore; both have been countered effectively.

The first claim—that Mendel failed to report crosses involving traits that did not show independent assortment—arises from the observation that all seven traits that Mendel studied do show independent assortment and that the pea plant has precisely seven pairs of chromosomes. For Mendel to have chosen seven genes, one located on each of the seven chromosomes, by chance alone seems extremely unlikely. In fact, the probability would be

$$7/7 \times 6/7 \times 5/7 \times 4/7 \times 3/7 \times 2/7 \times 1/7 = 0.006$$

That is, Mendel had less than one chance in one hundred of randomly picking seven traits on the seven different chromosomes. However, L. Douglas and E. Novitski in 1977 analyzed Mendel's data in a different way. To understand their analysis, you have to know that two genes sufficiently far apart on the same chromosome will appear to assort independently (to be discussed in chapter 6). Thus, Mendel's choice of characters showing independent assortment has to be viewed in light of the lengths of the chromosomes. That is, Mendel could have chosen two genes on the same chromosome that would still show independent assortment. In fact, he did. For example, stem length and pod texture (wrinkled or smooth) are on the fourth chromosome pair in peas. In their analysis, Douglas and Novitski report that the probability of randomly choosing seven characteristics that appear to assort independently is actually between one in four and one in three. So it seems that Mendel did not have to manipulate his choice of characters in order to hide the failure of independent assortment. He had a one in three chance of naively choosing the seven characters that he did, thereby uncovering no deviation from independent assortment.

The second claim—that Mendel fabricated data—comes from a careful analysis of Mendel's paper by R. A. Fisher, a brilliant English statistician and population geneticist. In a paper in 1936, Fisher pointed out two problems in Mendel's work. First, all of Mendel's published data taken together fit their expected ratios better than chance alone would predict. Second, some of Mendel's data fit incorrect expected ratios. This second "error" on Mendel's part came about as follows.

### Testcrossing Multihybrids

A simple test of Mendel's rule of independent assortment is the testcrossing of the dihybrid plant. We would predict, for example, that if we crossed an *RrYy* $F_1$ individual with an *rryy* individual, the results would include four phenotypes in a 1:1:1:1 ratio, as shown in figure 2.18. Mendel's data verified this prediction (box 2.2). We will proceed to look at a **trihybrid** cross in order to develop general rules for **multihybrids.**

A trihybrid Punnett square appears in figure 2.19. From this we can see that when a homozygous dominant and a homozygous recessive individual are crossed in the $P_1$ generation, plants in the $F_1$ generation are capable of producing eight gamete types. When these $F_1$ individuals are selfed, they in turn produce $F_2$ offspring of twenty-seven different genotypes in a ratio of sixty-fourths. By extrapolating from the monohybrid through the trihybrid, or simply by the rules of probability, we can construct table 2.3, which contains the rules for $F_1$ gamete production and $F_2$ zygote formation in a multihybrid cross. For example, from this table we can figure out the $F_2$ offspring when a dodecahybrid (twelve segregating genes: *AA BB CC . . . LL* × *aa bb cc . . . ll*) is selfed. The $F_1$ organisms in that cross will produce gametes with $2^{12}$, or 4,096, different genotypes. The proportion of homozygous recessive offspring in the $F_2$ generation is $1/(2^n)^2$ where $n = 12$, or 1 in 16,777,216. With complete dominance, there will be 4,096 different phenotypes in the $F_2$ generation. If dominance is incomplete, there can be $3^{12}$, or 531,441, different phenotypes in the $F_2$ generation.

## GENOTYPIC INTERACTIONS

Often, several genes contribute to the same phenotype. An example occurs in the combs of fowl (fig. 2.20). If we cross a rose-combed hen with a pea-combed rooster (or vice versa), all the $F_1$ offspring are walnut-combed. If we

Mendel determined whether a dominant phenotype in the $F_2$ generation was a homozygote or a heterozygote by self-fertilizing it and examining ten offspring. In an $F_2$ generation composed of 1*AA*:2*Aa*:1*aa,* he expected a 2:1 ratio of heterozygotes to homozygotes within the dominant phenotypic class. In fact, this ratio is not precisely correct because of the problem of misclassification of heterozygotes. It is probable that some heterozygotes will be classified as homozygotes because all their offspring will be of the dominant phenotype. The probability that one offspring from a selfed *Aa* individual has the dominant phenotype is 3/4, or 0.75: the probability that ten offspring will be of the dominant phenotype is $(0.75)^{10}$ or 0.056. Thus, Mendel misclassified heterozygotes as dominant homozygotes 5.6% of the time. He should have expected a 1.89:1.11 ratio instead of a 2:1 ratio to demonstrate segregation. Mendel classified 600 plants this way in one cross and got a ratio of 201 homozygous to 399 heterozygous offspring. This is an almost perfect fit to the presumed 2:1 ratio and thus a poorer fit to the real 1.89:1.11 ratio. This bias is consistent and repeated in Mendel's trihybrid analysis.

Fisher, believing in Mendel's basic honesty, suggested that Mendel's data do not represent an experiment but more of a hypothetical demonstration. In 1971, F. Weiling published a more convincing case in Mendel's defense. Pointing out that the data of Mendel's rediscoverers are also suspect for the same reason, he suggested that the problem lies with the process of pollen formation in plants, not with the experimenters. In an *Aa* heterozygote, two *A* and two *a* cells develop from a pollen mother cell. These cells tend to stay together on the anther. Thus, pollen cells do not fertilize in a strictly random fashion. A bee is more likely to take equal numbers of *A* and *a* pollen than chance alone would predict. The result is that the statistics Fisher used are not applicable. By using a different statistic, Weiling showed that, in fact, Mendel need not have manipulated any numbers (nor would have his rediscoverers) in order to get data that fit the expected ratios well. By the same reasoning, very little misclassification of heterozygotes would have occurred.

More recently, Weiling and others have made several additional points. First, for Mendel to be sure of ten offspring, he probably examined more than ten, and thus he probably kept his misclassification rate lower than 5.6%. Second, despite Fisher's brilliance as a statistician, several have made compelling arguments that Fisher's statistical analyses were incorrect. In other words, for subtle statistical reasons, many of his analyses involved methods and conclusions that were in error.

We conclude that there is no compelling evidence to suggest that Mendel in any way manipulated his data to demonstrate his rules. In fact, taking into account what is known about him personally, it is much more logical to believe that he did not "cheat."

cross the hens and roosters of this heterozygous $F_1$ group, we will get, in the $F_2$ generation, walnut-, rose-, pea-, and single-combed fowl in a ratio of 9:3:3:1. Can you figure out the genotypes of this $F_2$ population before reading further? An immediate indication that two allelic pairs are involved is the fact that the 9:3:3:1 ratio appeared in the $F_2$ generation. As we have seen, this ratio comes about when we cross dihybrids in which both genes have alleles that control traits with complete dominance.

Figure 2.21 shows the analysis of this cross. When dominant alleles of both genes are present in an individual (*R- P-*), the walnut comb appears. (The dash indicates any second allele; thus, *R- P-* could be *RRPP, RrPP, RRPp,* or *RrPp.*) A dominant allele of the rose gene (*R-*) with

**Table 2.3 Multihybrid Self-Fertilization, Where *n* Equals Number of Genes Segregating Two Alleles Each**

| | Monohybrid $n = 1$ | Dihybrid $n = 2$ | Trihybrid $n = 3$ | General Rule |
|---|---|---|---|---|
| Number of $F_1$ gametic genotypes | 2 | 4 | 8 | $2^n$ |
| Proportion of recessive homozygotes among the $F_2$ individuals | 1/4 | 1/16 | 1/64 | $1/(2^n)^2$ |
| Number of different $F_2$ phenotypes, given complete dominance | 2 | 4 | 8 | $2^n$ |
| Number of different genotypes (or phenotypes, if no dominance exists) | 3 | 9 | 27 | $3^n$ |

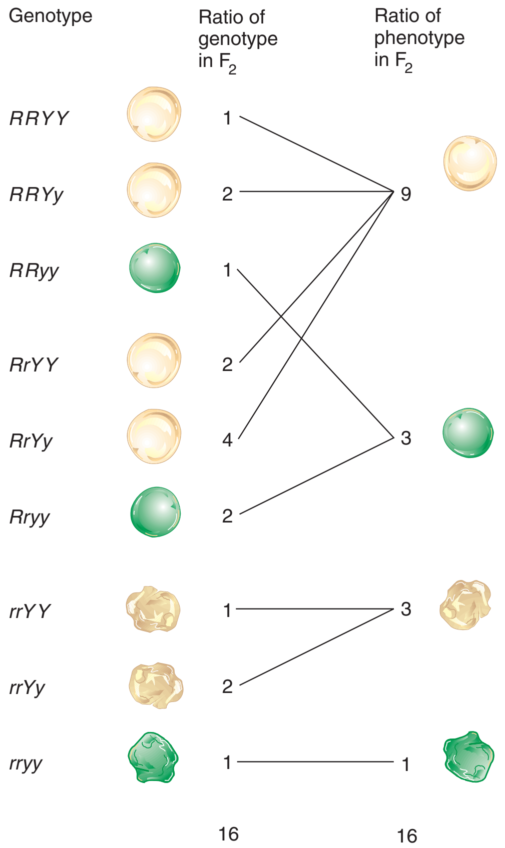


**Figure 2.16** The phenotypic and genotypic ratios of the offspring of dihybrid peas.

recessive alleles of the pea gene (*pp*) gives a rose comb. A dominant allele of the pea gene (*P*-) with recessive alleles of the rose gene (*rr*) gives pea-combed fowl. When both genes are homozygous for the recessive alleles, the fowl are single-combed. Thus, a 9:3:3:1 $F_2$ ratio arises from crossing dihybrid individuals even though different expressions of the same phenotypic characteristic, the comb, are involved. In our previous 9:3:3:1 example (see fig. 2.15), we dealt with two separate characteristics: shape and color of peas.

In corn (or maize, *Zea mays*), several different field varieties produce white kernels on the ears. In certain crosses, two white varieties will result in an $F_1$ generation with all purple kernels. If plants grown from these purple kernels are selfed, the $F_2$ individuals have both purple and white kernels in a ratio of 9:7. How can we explain this? We must be dealing with the offspring of dihybrids with each gene segregating two alleles, because the ratio is in sixteenths. Furthermore, we can see that the $F_2$ 9:7 ratio is a variation of the 9:3:3:1 ratio. The 3, 3, and 1 categories here are producing the same phenotype and thus make up 7/16 of the $F_2$ offspring. Figure 2.22 outlines the cross. We can see from this figure that the purple color appears only when dominant alleles of both genes are present. When one or both genes have only recessive alleles, the kernels will be white.

## Epistasis

The color of corn kernels illustrates the concept of **epistasis,** the interaction of nonallelic genes in the formation

**Figure 2.17** Independent assortment of two blood systems in human beings. In the ABO system, the $I^A$ and $I^B$ alleles are codominant. In a simplified view of the Rhesus system, the Rh$^+$ phenotype (*D* allele) is dominant to the Rh$^-$ phenotype (*d* allele).

**$P_1$** $I^AI^ADD$ X $I^BI^Bdd$

**$F_1$** $I^AI^BDd$
($F_1$ X $F_1$)

**$F_2$**

| Female \ Male | $I^AD$ | $I^Ad$ | $I^BD$ | $I^Bd$ |
|---|---|---|---|---|
| $I^AD$ | $I^AI^ADD$ | $I^AI^ADd$ | $I^AI^BDD$ | $I^AI^BDd$ |
| $I^Ad$ | $I^AI^ADd$ | $I^AI^Add$ | $I^AI^BDd$ | $I^AI^Bdd$ |
| $I^BD$ | $I^AI^BDD$ | $I^AI^BDd$ | $I^BI^BDD$ | $I^BI^BDd$ |
| $I^Bd$ | $I^AI^BDd$ | $I^AI^Bdd$ | $I^BI^BDd$ | $I^BI^Bdd$ |

| **$F_2$ Summary** | | | | | | |
|---|---|---|---|---|---|---|
| Phenotype | A Rh$^+$ | A Rh$^-$ | B Rh$^+$ | B Rh$^-$ | AB Rh$^+$ | AB Rh$^-$ |
| Frequency | 3 | 1 | 3 | 1 | 6 | 2 |

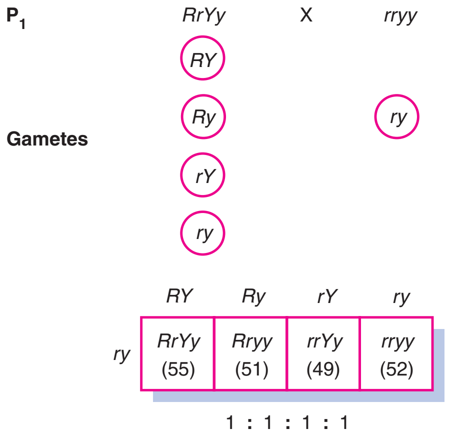


**Figure 2.18** Testcross of a dihybrid. A 1:1:1:1 ratio is expected in the offspring.

of the phenotype. This is a process analogous to dominance among alleles of one gene. For example, the recessive apterous (wingless) gene in fruit flies is epistatic to any gene that controls wing characteristics; hairy wing is hypostatic to apterous (that is, the recessive apterous gene, when homozygous, masks the presence of the hairy wing gene, because, obviously, without wings, no wing characteristics can be expressed). Note that the genetic control of comb type in fowl does not involve epistasis. There are no allelic combinations at one locus that mask genotypes at another locus: the 9:3:3:1 ratio is not an indication of epistasis. To illustrate further the principle of epistasis, we can look at the control of coat color in mice.

In one particular example, a pure-breeding black mouse is crossed with a pure-breeding albino mouse (pure white because all pigment is lacking); all of the offspring are agouti (the typical brownish-gray mouse color). When the F$_1$ agouti mice are crossed with each other, agouti, black, and albino offspring appear in the F$_2$ generation in a ratio of 9:3:4. What are the genotypes in this cross? The answer appears in figure 2.23. By now it should be apparent that the F$_2$ ratio of 9:3:4 is also a variant of the 9:3:3:1 ratio; it indicates epistasis in a dihybrid cross. What is the mechanism producing this 9:3:4 ratio? Of a potential 9:3:3:1 ratio, one of the 3/16 classes and the 1/16 class are combined to create a 4/16 class. Any genotype that includes $c^ac^a$ will be albino, masking the A gene, but as long as at least one dominant *C* allele is present, the A gene can express itself. Mice with dominant alleles of both genes (*A*-*C*-) will have the agouti color, whereas mice that are homozygous recessive at the A gene (*aaC*-) will be black. So, at the A gene, *A* for agouti

P$_1$ *AA BB CC* × *aa bb cc*

F$_1$ *Aa Bb Cc* × Self

| | *A B C* | *A B c* | *A b C* | *A b c* | *a B C* | *a B c* | *a b C* | *a b c* |
|---|---|---|---|---|---|---|---|---|
| *A B C* | *AA BB CC* | *AA BB Cc* | *AA Bb CC* | *AA Bb Cc* | *Aa BB CC* | *Aa BB Cc* | *Aa Bb CC* | *Aa Bb Cc* |
| *A B c* | *AA BB Cc* | *AA BB cc* | *AA Bb Cc* | *AA Bb cc* | *Aa BB Cc* | *Aa BB cc* | *Aa Bb Cc* | *Aa Bb cc* |
| *A b C* | *AA Bb CC* | *AA Bb Cc* | *AA bb CC* | *AA bb Cc* | *Aa Bb CC* | *Aa Bb Cc* | *Aa bb CC* | *Aa bb Cc* |
| *A b c* | *AA Bb Cc* | *AA Bb cc* | *AA bb Cc* | *AA bb cc* | *Aa Bb Cc* | *Aa Bb cc* | *Aa bb Cc* | *Aa bb cc* |
| *a B C* | *Aa BB CC* | *Aa BB Cc* | *Aa Bb CC* | *Aa Bb Cc* | *aa BB CC* | *aa BB Cc* | *aa Bb CC* | *aa Bb Cc* |
| *a B c* | *Aa BB Cc* | *Aa BB cc* | *Aa Bb Cc* | *Aa Bb cc* | *aa BB Cc* | *aa BB cc* | *aa Bb Cc* | *aa Bb cc* |
| *a b C* | *Aa Bb CC* | *Aa Bb Cc* | *Aa bb CC* | *Aa bb Cc* | *aa Bb CC* | *aa Bb Cc* | *aa bb CC* | *aa bb Cc* |
| *a b c* | *Aa Bb Cc* | *Aa Bb cc* | *Aa bb Cc* | *Aa bb cc* | *aa Bb Cc* | *aa Bb cc* | *aa bb Cc* | *aa bb cc* |

**Figure 2.19** Trihybrid cross.

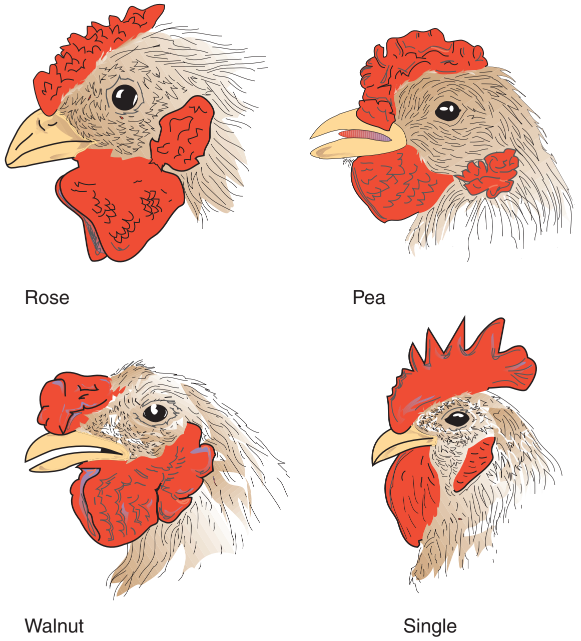
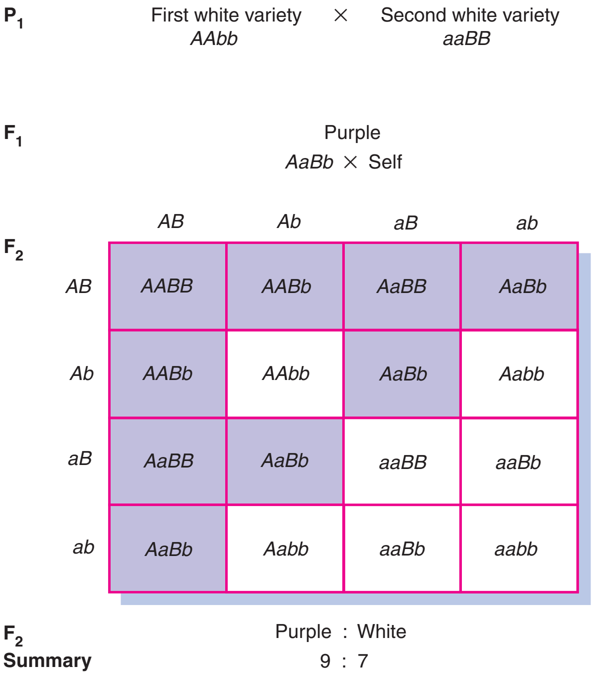


**Figure 2.20** Four types of combs in fowl.



$P_1$ First white variety *AAbb* × Second white variety *aaBB*

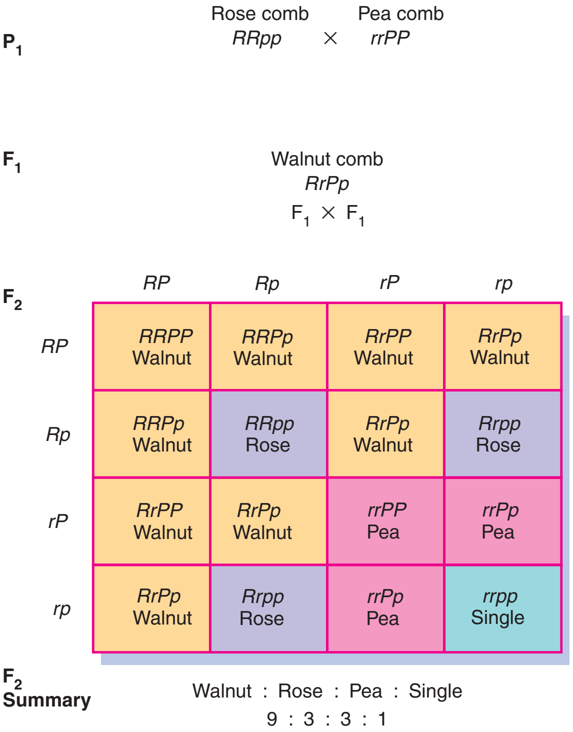
$F_1$ Purple *AaBb* × Self

$F_2$

| | *AB* | *Ab* | *aB* | *ab* |
|---|---|---|---|---|
| *AB* | *AABB* | *AABb* | *AaBB* | *AaBb* |
| *Ab* | *AABb* | *AAbb* | *AaBb* | *Aabb* |
| *aB* | *AaBB* | *AaBb* | *aaBB* | *aaBb* |
| *ab* | *AaBb* | *Aabb* | *aaBb* | *aabb* |

$F_2$ Summary Purple : White 9 : 7

**Figure 2.22** Color production in corn.



$P_1$ Rose comb *RRpp* × Pea comb *rrPP*

$F_1$ Walnut comb *RrPp* $F_1$ × $F_1$

$F_2$

| | *RP* | *Rp* | *rP* | *rp* |
|---|---|---|---|---|
| *RP* | *RRPP*<br>Walnut | *RRPp*<br>Walnut | *RrPP*<br>Walnut | *RrPp*<br>Walnut |
| *Rp* | *RRPp*<br>Walnut | *RRpp*<br>Rose | *RrPp*<br>Walnut | *Rrpp*<br>Rose |
| *rP* | *RrPP*<br>Walnut | *RrPp*<br>Walnut | *rrPP*<br>Pea | *rrPp*<br>Pea |
| *rp* | *RrPp*<br>Walnut | *Rrpp*<br>Rose | *rrPp*<br>Pea | *rrpp*<br>Single |

$F_2$ Summary Walnut : Rose : Pea : Single 9 : 3 : 3 : 1

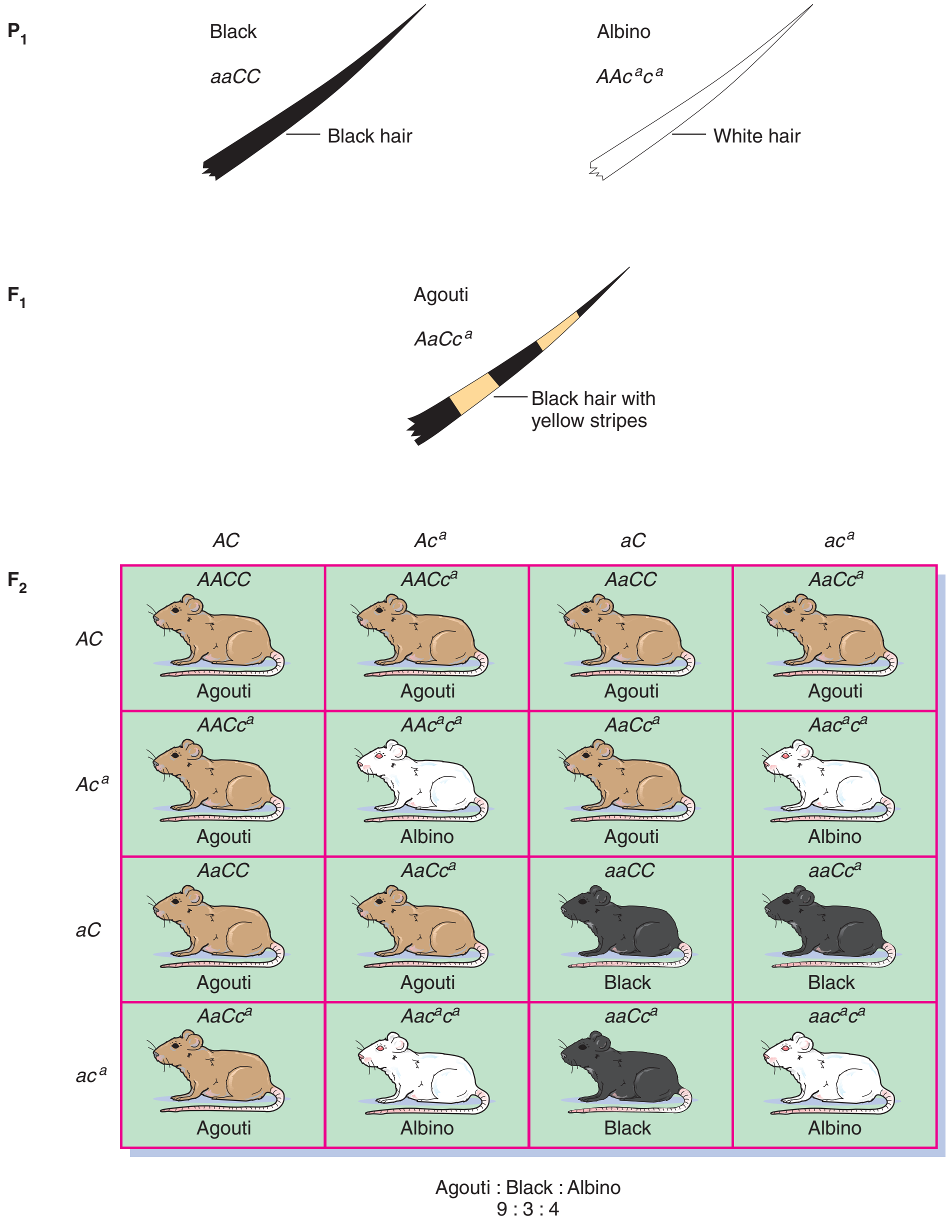
**Figure 2.21** Independent assortment in the determination of comb type in fowl.

is dominant to *a* for black. The albino gene ($c^a$), when homozygous, is epistatic to the A gene; the A gene is hypostatic to the gene for albinism.

## Mechanism of Epistasis

In this case, the physiological mechanism of epistasis is known. The pigment melanin is present in both the black and agouti phenotypes. The agouti is a modified black hair in which yellow stripes (the pigment phaeomelanin) have been added. Thus, with melanin present, agouti is dominant. Without melanin, we get an albino regardless of the genotype of the agouti gene because both agouti and black depend on melanin. Albinism is the result of one of several defects in the enzymatic pathway for the synthesis of melanin (fig. 2.24).

Knowing that epistatic modifications of the 9:3:3:1 ratio come about through gene interactions at the biochemical level, we can look for a biochemical explanation for the 9:7 ratio in corn kernel color (fig. 2.22). Two possible mechanisms for a 9:7 ratio are shown in figure 2.25. Either a two-step process takes a precursor molecule and turns it into purple pigment, or two precursors that must be converted to final products then combine to produce purple pigment. The dominant alleles from the two genes control the two steps in the process. Recessive alleles are ineffective. Thus, dominants are necessary for both steps to complete the pathways for a purple pig-



| | $AC$ | $Ac^a$ | $aC$ | $ac^a$ |
|---|---|---|---|---|
| $AC$ | $AACC$<br>Agouti | $AACc^a$<br>Agouti | $AaCC$<br>Agouti | $AaCc^a$<br>Agouti |
| $Ac^a$ | $AACc^a$<br>Agouti | $AAc^ac^a$<br>Albino | $AaCc^a$<br>Agouti | $Aac^ac^a$<br>Albino |
| $aC$ | $AaCC$<br>Agouti | $AaCc^a$<br>Agouti | $aaCC$<br>Black | $aaCc^a$<br>Black |
| $ac^a$ | $AaCc^a$<br>Agouti | $Aac^ac^a$<br>Albino | $aaCc^a$<br>Black | $aac^ac^a$<br>Albino |

**Figure 2.23** Epistasis in the coat color of mice.

ment. Stopping the process at any point prevents the production of purple color.

Another example of epistasis occurs in the snapdragon (*Antirrhinum majus*). There, a gene called *nivea* has alleles that determine whether any pigment is produced; the *nn* genotype prevents pigment production, whereas the *NN* or *Nn* genotypes permit pigment color genes to express themselves. The *eosinea* gene controls the production of a red anthocyanin pigment. In the presence of the *N* allele of the *nivea* gene, the genotypes *EE* or *Ee* of the *eosinea* gene produce red flowers; the *ee* genotype produces pink flowers. When dihybrids are self-fertilized, red-, pink-, and white-flowered plants are produced in a ratio of 9:3:4 (fig. 2.26). The epistatic interaction is the *nn* genotype masking the expression of alleles at the *eosinea* gene. In other words, regardless of the genotypes of the *eosinea* gene (*EE, Ee,* or *ee*), the flowers will be white if the *nivea* gene has the *nn*

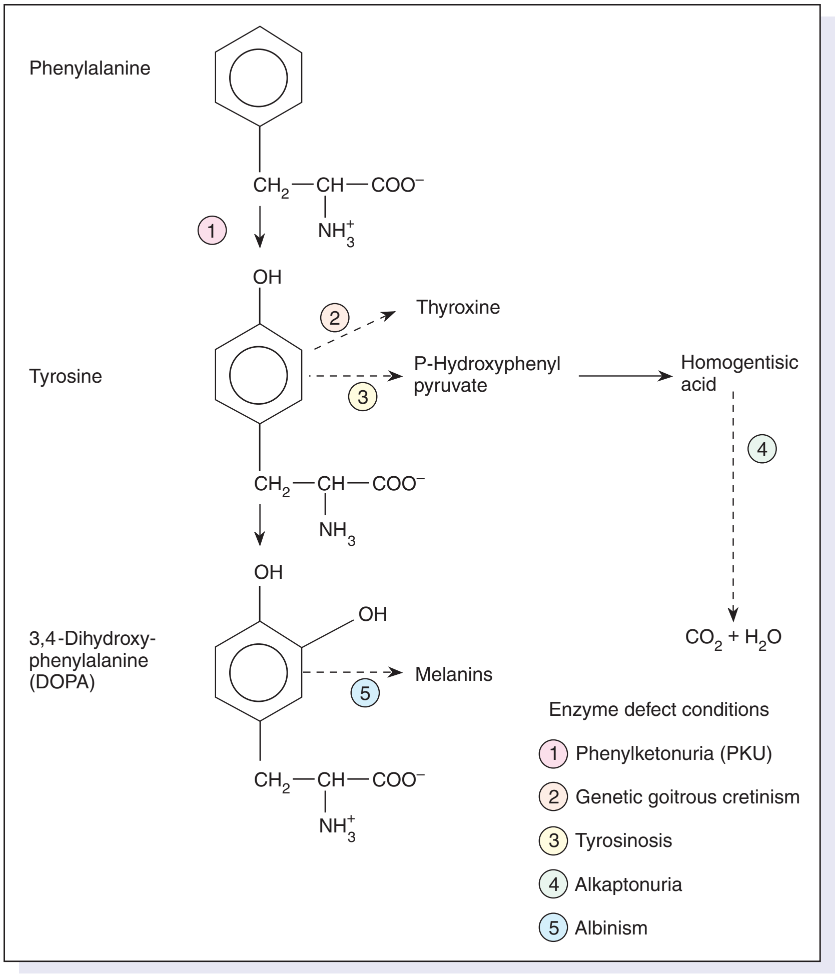

Phenylalanine
CH2—CH—COO⁻
NH3+
1
OH
2
Thyroxine
Tyrosine
3
P-Hydroxyphenyl pyruvate
Homogentisic acid
4
CH2—CH—COO⁻
NH3
OH
OH
3,4-Dihydroxy-phenylalanine (DOPA)
5
Melanins
CO2 + H2O
CH2—CH—COO⁻
NH3+
Enzyme defect conditions
1 Phenylketonuria (PKU)
2 Genetic goitrous cretinism
3 Tyrosinosis
4 Alkaptonuria
5 Albinism

**Figure 2.24** In humans, errors in melanin synthesis produce different physical conditions and diseases, depending on which part of the tyrosine (an amino acid) metabolic pathway is disrupted. The broken arrows indicate that there is more than one step in the pathways; the conditions listed occur only in homozygous recessives.

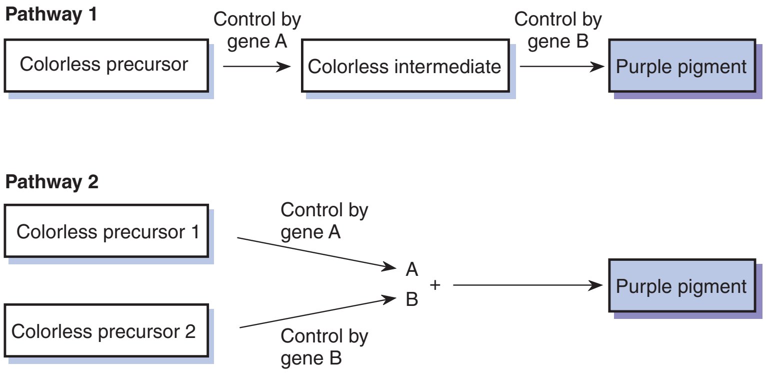

Pathway 1
Colorless precursor
Control by gene A
Colorless intermediate
Control by gene B
Purple pigment
Pathway 2
Colorless precursor 1
Control by gene A
A
+
B
Colorless precursor 2
Control by gene B
Purple pigment

**Figure 2.25** Possible metabolic pathways of color production that would yield 9:7 ratios in the $F_2$ generation of a self-fertilized dihybrid.

combination of alleles. Thus, *nivea* is epistatic to *eosinea,* and *eosinea* is hypostatic to *nivea.* (We should add that at least seven major colors occur in snapdragons, along with subtle shade differences, all genetically controlled by the interactions of at least seven genes.)

Other types of epistatic interactions occur in other organisms. Table 2.4 lists several. We do not know the exact physiological mechanisms in many cases, especially when developmental processes are involved (e.g., size and shape). However, from an analysis of crosses, we can know the number of genes involved and the general nature of their interactions.

## BIOCHEMICAL GENETICS

### Inborn Errors of Metabolism

The examples of mouse coat color, corn kernel color, and snapdragon flower petal color demonstrate that genes control the formation of enzymes, proteins that control the steps in biochemical pathways. For the most part, dominant alleles control functioning enzymes that catalyze biochemical steps. Recessive alleles often produce nonfunctioning enzymes that cannot catalyze specific steps. Often a heterozygote is normal because one allele produces a functional enzyme; usually only half the enzyme quantity of the dominant homozygote is enough. The study of the relationship between genes and enzymes is generally called **biochemical genetics** because it involves the genetic control of biochemical pathways. A. E. Garrod, a British physician, pointed out this general concept of human gene action in *Inborn Errors of Metabolism,* published in 1909. Only nine years after Mendel was rediscovered, Garrod described several human conditions, such as albinism and alkaptonuria, that occur in individuals who are homozygous for recessive alleles (see fig. 2.24).

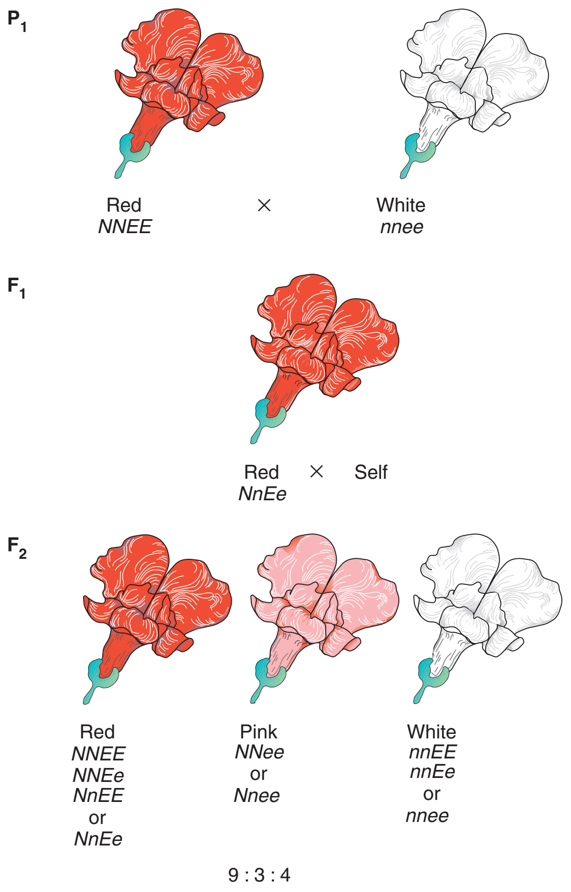


**Figure 2.26** Flower color inheritance in snapdragons. This is an example of epistasis: an *nn* genotype masks the expression of alleles (*EE, Ee,* or *ee*) at the *eosinea* gene.

**Table 2.4** **Some Examples of Epistatic Interactions Among Alleles of Two Genes**

| Characteristic | Phenotype of F$_1$ Dihybrid (*AaBb*) | Phenotypic F$_2$ Ratio |
|---|---|---|
| Corn and sweet pea color | Purple | Purple:white<br>9:7 |
| Mouse coat color | Agouti | Agouti:black:albino<br>9:3:4 |
| Shepherd's purse seed capsule shape | Triangular | Triangular:oval<br>15:1 |
| Summer squash shape | Disk | Disk:sphere:elongate<br>9:6:1 |
| Fowl color | White | White:colored<br>13:3 |

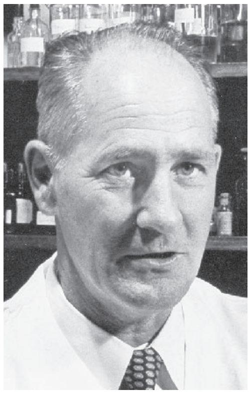
For example, people normally degrade homogentisic acid (alkapton) into maleylacetoacetic acid. Persons with the disease alkaptonuria are homozygous for a nonfunctional form of the enzyme essential to the process: homogentisic acid oxidase, found in the liver. Absence of this enzyme blocks the degradation reaction so that homogentisic acid builds up. This acid darkens upon oxidation. Thus, affected persons can be identified by the black color of their urine after its exposure to air. Eventually, alkaptonuria causes problems in the joints and a darkening of cartilage that is visible in the ears and the eye sclera.

## *One-Gene-One-Enzyme Hypothesis*

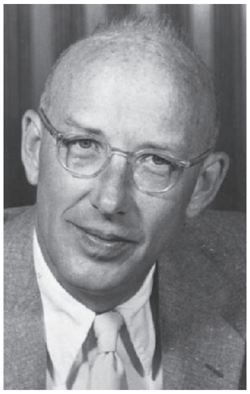
Pioneering work in the concept that genes control the production of enzymes, which in turn control the steps in biochemical pathways, was done by George Beadle and Edward Tatum, who eventually shared the Nobel Prize for their work. They not only put forth the **one-gene-one-enzyme hypothesis,** but also used mutants to work out the details of biochemical pathways. In 1941, Beadle and Tatum were the first scientists to isolate mutants with nutritional requirements that defined steps in a biochemical pathway. In the early 1940s, they united the fields of biochemistry and genetics by using strains of a bread mold with specific nutritional requirements to discover the steps in biochemical pathways in that organism.

Through this century, the study of mutations has been the driving force in genetics. The process of mutation produces alleles that differ from the wild-type and shows us that a particular aspect of the phenotype is under genetic control. Beadle and Tatum used mutants to work out the steps in the biosynthesis of niacin (vitamin $B_3$) in pink bread mold, *Neurospora crassa.*

Normally, *Neurospora* synthesizes niacin via the pathway shown in figure 2.27. Beadle and Tatum isolated mutants that could not grow unless niacin was provided in the culture medium; these mutants had enzyme deficiencies in the synthesizing pathway that ends with niacin. Thus, although wild-type *Neurospora* could grow on a medium without additives, the mutants could not. Beadle and Tatum had a general idea, based on the structure of niacin, as to what substances would be in the niacin biosynthesis pathway. They could thus make educated guesses as to what substances they might add to the culture medium to enable the mutants to grow. Mutant B (table 2.5), for example, could grow if given niacin or, alternatively, 3-hydroxyanthranilic acid. It could not grow if given only kynurenine. Thus, Beadle and Tatum knew that the B mutation affected the pathway between kynurenine and 3-hydroxyanthranilic acid. Similarly, mutant A could grow if given 3-hydroxyanthranilic acid or kynurenine instead of niacin. Therefore, these two prod-



George W. Beadle (1903–89). Courtesy of the Archives, California Institute of Technology.



Edward L. Tatum (1909–75). Courtesy of the Proceedings for the National Academy of Sciences.

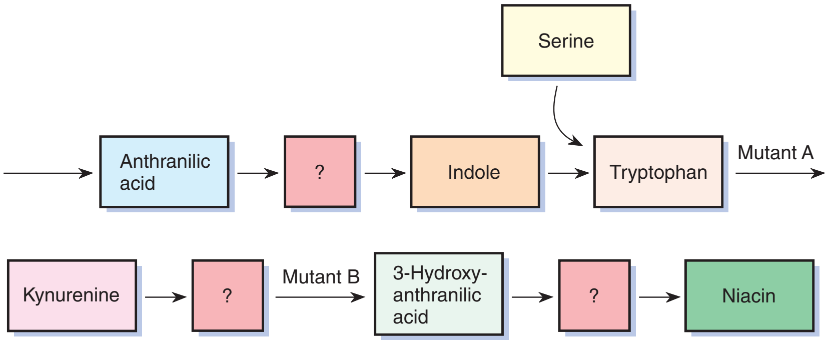


**Figure 2.27** Pathway of niacin synthesis in *Neurospora.* Each *arrow* represents an enzyme-mediated step. Each *question mark* represents a presumed but (at the time Beadle and Tatum were working) unknown compound.

**Table 2.5** **Growth Performance of *Neurospora* Mutants (plus sign indicates growth; minus sign indicates no growth)**

| | Additive | | | |
|---|---|---|---|---|
| | **Tryptophan** | **Kynurenine** | **3-Hydroxyanthranilic Acid** | **Niacin** |
| Wild-type | + | + | + | + |
| Mutant A | − | + | + | + |
| Mutant B | − | − | + | + |

ucts must be in the pathway *after* the step interrupted in mutant A. Conversely, since neither of these mutant organisms could grow when given only tryptophan, Beadle and Tatum knew that tryptophan occurred in the pathway before the steps with the deficient enzymes. By this type of analysis, they discovered the steps in several biochemical pathways of *Neurospora.* Many biochemical pathways are similar in a huge range of organisms, and thus Beadle and Tatum's work was of general importance. (We will spend more time studying *Neurospora* in chapter 6.)

Beadle and Tatum could further verify their work by observing which substances accumulated in the mutant organisms. If a biochemical pathway is blocked at a certain point, then the substrate at that point cannot convert into the next product, and it builds up in the cell. For example, in the niacin pathway (fig. 2.27), if a block occurs just after 3-hydroxyanthranilic acid, that substance will build up in the cell because it cannot convert into the next substance on the way to niacin.

This analysis could be misleading, however, if the built-up substance is being "siphoned off" into other biochemical pathways in the cell. Also, the cell might attempt to break down or sequester toxic substances. This would mean there might not be an obvious buildup of the substance just before the blocked step.

Beadle and Tatum concluded from their studies that one gene controls the production of one enzyme. The one-gene-one-enzyme hypothesis is an oversimplification that we will clarify later in the book. As a rule of thumb, however, the hypothesis is valid, and it has served to direct attention to the functional relationship between genes and enzymes in biochemical pathways.

Although a change in a single enzyme usually disrupts a single biochemical pathway, it frequently has more than one effect on phenotype. Multiple effects are referred to as **pleiotropy.** A well-known example is sickle-cell anemia, caused by a mutation in the gene for the β chain of the hemoglobin molecule. In a homozygote, this mutation causes a sickling of red blood cells (fig. 2.28). The sickling of these cells has two major ramifications.

First, the liver destroys the sickled cells, causing anemia. The phenotypic effects of this anemia include physical weakness, slow development, and hypertrophy of the bone marrow, resulting in the "tower skull" seen in some of those afflicted with the disease. The second major effect of sickle-cell anemia is that the sickled cells interfere with capillary blood flow, clumping together and resulting in damage to every major organ. The individual can suffer pain, heart failure, rheumatism, and other ill effects. Hence, a single mutation shows itself in many aspects of the phenotype.

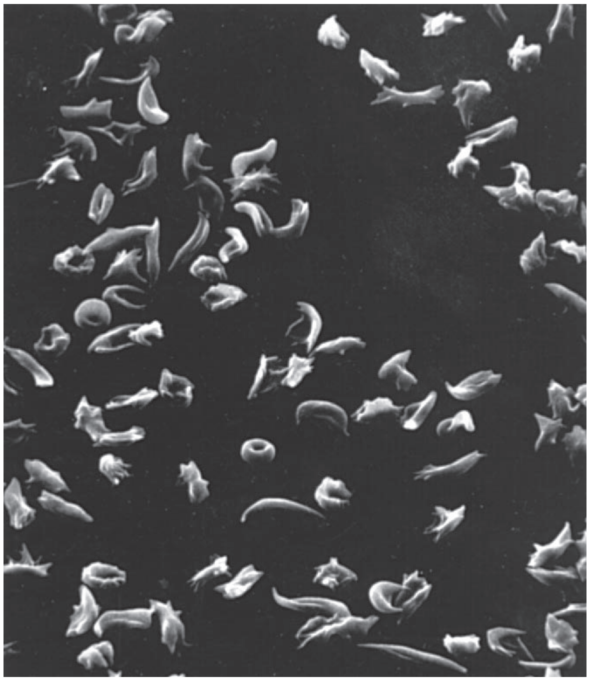


**Figure 2.28** Sickle-shaped red blood cells from a person with sickle-cell anemia. Red blood cells are about 7 to 8 μm in diameter. (Courtesy of Dr. Patricia N. Farnsworth.)

## SUMMARY

**STUDY OBJECTIVE 1:** To understand that genes are discrete units that control the appearance of an organism 17–18

Genes control phenotypic traits such as size and color. They are inherited as discrete units.

**STUDY OBJECTIVE 2:** To understand Mendel's rules of inheritance: segregation and independent assortment 18–22

Higher organisms contain two alleles of each gene, but only one allele enters each gamete. Zygote formation restores the double number of alleles in the cell. This is Mendel's rule of segregation.

Alleles of different genes segregate independently of each other. Mendel was the first to recognize the 3:1 phenotypic ratio as a pattern of inheritance; the 9:3:3:1 ratio demonstrates independent assortment in hybrids. Mendel was successful in his endeavor because he performed careful experiments using discrete characteristics, large numbers of offspring, and an organism (the pea plant) amenable to controlled fertilizations.

**STUDY OBJECTIVE 3:** To understand that dominance is a function of the interaction of alleles; similarly, epistasis is a function of the interaction of nonallelic genes 22–37

There can be many alleles for one gene, although each individual organism has only two alleles for each gene. A phenotype is dominant if it is expressed when one or two copies of its allele are present (heterozygote or homozygote). Dominance depends, however, on the level of the phenotype one looks at.

Genes usually control the production of enzymes, which control steps in metabolic pathways. Many human metabolic diseases are due to homozygosity of an allele that produces a nonfunctioning enzyme.

Nonallelic genes can interact in producing a phenotype so that alleles of one gene mask the expression of alleles of another gene. This process, termed *epistasis,* alters the expected phenotypic ratios.

**STUDY OBJECTIVE 4:** To define how genes generally control the production of enzymes and thus the fate of biochemical pathways 37–39

Beadle and Tatum used mutants with mutations in the niacin biosynthesis pathway to work out the steps in the pathway. A single mutation can have many phenotypic effects (pleiotropy).

## SOLVED PROBLEMS

**PROBLEM 1:** In corn, rough sheath (*rs*) is recessive to smooth sheath (*Rs*), midrib absent (*mrl*) is recessive to midrib present (*Mrl*), and crinkled leaf (*cr*) is recessive to smooth leaf (*Cr*). (Alleles are named for the mutants, which are all recessive.) What are the results of testcrossing a trihybrid?

***Answer:*** The trihybrid has the genotype *Rsrs Mrlmrl Crcr.* This parent is capable of producing eight different gamete types in equal frequencies, all combinations of one allele from each gene (*Rs Mrl Cr, Rs Mrl cr, Rs mrl Cr, Rs mrl cr, rs Mrl Cr, rs Mrl cr, rs mrl Cr,* and *rs mrl cr*). In a testcross, the other parent is a recessive homozygote with the genotype *rsrs mrlmrl crcr,* capable of producing only one type of gamete, with the alleles *rs mrl cr.* Thus, this cross can produce zygotes of eight different genotypes (and phenotypes), one for each of the gamete types of the trihybrid parent: *Rsrs Mrlmrl Crcr* (smooth sheath, midrib present, smooth leaf); *Rsrs Mrlmrl crcr* (smooth sheath, midrib present, crinkled leaf); *Rsrs mrlmrl Crcr* (smooth sheath, midrib absent, smooth leaf); *Rsrs mrlmrl crcr* (smooth sheath, midrib absent, crinkled leaf); *rsrs Mrlmrl Crcr* (rough sheath, midrib present, smooth leaf); *rsrs Mrlmrl crcr* (rough sheath, midrib present, crinkled leaf); *rsrs mrlmrl Crcr* (rough sheath, midrib absent, smooth leaf); and *rsrs mrlmrl crcr* (rough sheath, midrib absent, crinkled leaf). Each should make up one-eighth of the total number of offspring.

**PROBLEM 2:** Summer squash come in three shapes: disk, spherical, and elongate. In one experiment, researchers crossed two squash plants with disk-shaped fruits. The first 160 seeds planted from this cross produced plants with fruit shapes as follows: 89 disk, 61 sphere, and 10 elongate. What is the mode of inheritance of fruit shape in summer squash?

***Answer:*** The numbers are very close to a ratio of 90:60:10, or 9:6:1, an epistatic variant of the 9:3:3:1, with the two 3/16ths categories combined. If this is the case, then the parent plants with disk-shaped fruits were dihybrids (*AaBb*). Among the offspring, 9/16ths had disk-shaped fruit, indicating that it takes at least one dominant allele of each gene to produce disk-shaped fruits (*A-B-*: *AABB, AaBB, AABb,* or *AaBb*). The 1/16th category of plants with elongate fruits indicates that this fruit shape occurs in homozygous recessive plants (*aabb*). The plants with spherical fruit are thus plants with a dominant allele of one gene but a homozygous recessive combination at the other gene (*AAbb, Aabb, aaBB,* or *aaBb*). In summary, then, two genes combine to control fruit shape in summer squash. The epistatic interactions between the two genes produce a 9:6:1 ratio of offspring phenotypes when dihybrids are crossed.

**PROBLEM 3:** A geneticist studying the pathway of synthesis of phenylalanine in *Neurospora* isolated several mutants that require phenylalanine to grow. She tested whether each mutant would grow when provided additives that she believed were in the pathway of phenylalanine synthesis (see table); a plus indicates growth and minus indicates the lack of growth in the three mutants tested.

| | **Additive** | | | |
|---|---|---|---|---|
| | **Phenylpyruvate** | **Prephenate** | **Chorismate** | **Phenylalanine** |
| Wild-type | + | + | + | + |
| Mutant 1 | − | − | − | + |
| Mutant 2 | + | + | − | + |
| Mutant 3 | + | − | − | + |

Where in the pathway to phenylalanine synthesis does each of the additives belong, if at all?

***Answer:*** The wild-type grows in the presence of all additives. This is not surprising since the wild-type can grow, by definition, in the absence of all the additives because it can synthesize phenylalanine *de novo.* Mutant 1 cannot grow in the presence of any additive except phenylalanine, indicating that its mutation affects the step just before the end of the pathway at phenylalanine. In other words, each of the other additives occurs in the phenylalanine pathway before the point of the mutation in mutant 1. Mutant 2 can grow if given any additive but chorismate, indicating that chorismate is at the beginning of the pathway, and the mutation affects the pathway just after that step. Finally, mutant 3 can grow if given phenylpyruvate or phenylalanine, indicating that its mutation affects the step before phenylpyruvate and phenylalanine, but after the earlier part of the pathway. Putting all of this information together indicates that the pathway to phenylalanine, with mutants indicated, is:

$$\text{chorismate} \xrightarrow{2} \text{prephenate} \xrightarrow{3} \text{phenylpyruvate} \xrightarrow{1} \text{phenylalanine}$$

## EXERCISES AND PROBLEMS*

### SEGREGATION

**1.** Mendel crossed tall pea plants with dwarf ones. The $F_1$ plants were all tall. When these $F_1$ plants were selfed to produce the $F_2$ generation, he got a 3:1 tall-to-dwarf ratio in the offspring. Predict the genotypes and phenotypes and relative proportions of the $F_3$ generation produced when the $F_2$ generation was selfed.

**2.** What properties of fruit flies and corn made them the organisms of choice for geneticists during most of the first half of the twentieth century? (Molecular geneticists have made great strides working with bacteria and viruses. You could begin thinking at this point about the properties that have made these organisms so valuable to geneticists.)

**3.** State precisely the rules of segregation and independent assortment. (See also the Exercises and Problems section on Independent Assortment.)

**4.** In *Drosophila,* a cross between a dark-bodied fly and a tan-bodied fly yields seventy-six tan and eighty dark flies. Diagram the cross.

**5.** If two black mice are crossed, ten black and three white mice result.
   **a.** Which allele is dominant?
   **b.** Which allele is recessive?
   **c.** What are the genotypes of the parents?

**6.** In *Drosophila,* two red-eyed flies mate and yield 110 red-eyed and 35 brown-eyed offspring. Diagram the cross and determine which allele is dominant.

### DOMINANCE IS NOT UNIVERSAL

**7.** Explain how Tay-Sachs disease can be both a recessive and an incomplete dominant trait. What are the differences between incomplete dominance and codominance?

**8.** How does the biochemical pathway in figure 2.13 explain how alleles $I^A$ and $I^B$ are codominant, yet both dominant to allele *i*?

---

*Answers to selected exercises and problems are on page A-1.

**9.** Two short-eared pigs are mated. In the progeny, three have no ears, seven have short ears, and four have long ears. Explain these results by diagramming the cross.

**10.** A plant with red flowers is crossed with a plant with white flowers. All the progeny are pink. When the plants with pink flowers are crossed, the progeny are eleven red, twenty-three pink, and twelve white. What is the mode of inheritance of flower color?

### NOMENCLATURE

**11.** In fruit flies, a new dominant trait, washed eye, was discovered. Describe different ways of naming the alleles of the washed-eye gene.

**12.** The following is a list of ten genes in fruit flies, each with one of its alleles given. Are the alleles shown dominant or recessive? Are they mutant or wild-type? What is the alternative allele for each? Is the alternative allele dominant or recessive in each case?

| Name of Gene | Allele |
|---|---|
| *yellow* | $y^+$ |
| *Hairy wing* | *Hw* |
| *Abruptex* | $Ax^+$ |
| *Confluens* | *Co* |
| *raven* | $rv^+$ |
| *downy* | *dow* |
| *Minute(2)e* | $M(2)e^+$ |
| *Jammed* | *J* |
| *tufted* | $tuf^+$ |
| *burgundy* | *bur* |

### MULTIPLE ALLELES

**13.** In the ABO blood system in human beings, alleles $I^A$ and $I^B$ are codominant, and both are dominant to the *i* allele. In a paternity dispute, a type AB woman claimed that one of four men, each with different blood types, was the father of her type A child. Which of the following could be the blood type of the father of the child on the basis of the evidence given?

**a.** Type A
**b.** Type B
**c.** Type O
**d.** Type AB

**14.** Under what circumstances can the phenotypes of the ABO system be used to refute paternity?

**15.** In blood transfusions, one blood type is called the "universal donor" and one the "universal recipient" because of their ABO compatibilities. Which is which?

**16.** Among the genes having the greatest number of alleles are those involved in self-incompatibility in plants. In some cases, hundreds of alleles exist for a single gene. What types of constraints might exist to set a limit on the number of alleles a gene can have?

**17.** In the human ABO blood system, the alleles $I^A$ and $I^B$ are dominant to *i.* What possible phenotypic ratios do you expect from a mating between a type A individual and a type B individual?

**18.** In screech owls, crosses between red and silver individuals sometimes yield all red; sometimes 1/2 red:1/2 silver; and sometimes 1/2 red:1/4 white:1/4 silver offspring. Crosses between two red owls yield either all red, 3/4 red:1/4 silver, or 3/4 red:1/4 white offspring. What is the mode of inheritance?

**19.** A premed student, Steve, plans to marry the daughter of the dean of nursing. The dean's husband was sterile, and the daughter was conceived by artificial insemination. Steve's father puts pressure on Steve to marry someone else. Having served as an anonymous sperm donor, he is concerned that Steve and his fiancé may be half brother and sister. Given the following information, deduce whether Steve and his fiancé could be related. (The MN and Ss systems are two independent, codominant blood-type systems.)

| | Blood Type |
|---|---|
| Dean | A, MN, Ss |
| Her daughter | O, M, S |
| Steve's father | A, MN, Ss |
| Steve | O, N, s |
| Steve's mother | B, N, s |

### INDEPENDENT ASSORTMENT

**20.** Mendel self-fertilized dihybrid plants (*RrYy*) with round and yellow seeds and got a 9:3:3:1 ratio in the $F_2$ generation. As a test of Mendel's hypothesis of independent assortment, predict the kinds and numbers of progeny produced in testcrosses of these $F_2$ offspring.

**21.** Four o'clock plants have a gene for color and a gene for height with the following phenotypes:

| | |
|---|---|
| *RR:* red flower | *TT:* tall plant |
| *Rr:* pink flower | *Tt:* medium height plant |
| *rr:* white flower | *tt:* dwarf plant |

Give the proportions of genotypes and phenotypes produced if a dihybrid plant is self-fertilized.

**22.** A particular variety of corn has a gene for kernel color and a gene for height with the following phenotypes:

| | |
|---|---|
| *CC, Cc:* purple kernels | *TT:* tall stem |
| *cc:* white kernels | *Tt:* medium height stem |
| | *tt:* dwarf stem |

Give the proportions of genotypes and phenotypes produced if a dihybrid plant is selfed.

**23.** To determine the genotypes of the offspring of a cross in which a corn trihybrid (*Aa Bb Cc*) was selfed, a geneticist has three choices. He or she can take a sample of the progeny and (a) self-fertilize the individual plants, (b) testcross the plants, or (c) cross the individuals with a trihybrid (backcross). Which method is preferable?

**24.** In figure 2.17, the $F_2$ phenotypic ratio is 3:1:3:1:6:2. What are the phenotypic segregation ratios for each blood system (AB, Rh) separately? Are they segregating properly? What phenotypic ratio in the $F_2$ generation would indicate interference with independent assortment?

**25.** Assume that Mendel looked simultaneously at four traits of his pea plants (and each trait exhibited dominance). If he crossed a homozygous dominant plant with a homozygous recessive plant, all the $F_1$ offspring would be of the dominant phenotype. If he then selfed the $F_1$ plants, how many different types of gametes would these $F_1$ plants produce? How many different phenotypes would appear in the $F_2$ generation? How many different genotypes would appear? What proportion of the $F_2$ offspring would be of the fourfold recessive phenotype?

**26.** A geneticist crossed two corn plants, creating an $F_1$ decahybrid (ten segregating loci). He then self-fertilized this decahybrid. How many different kinds of gametes did the $F_1$ plant produce? What proportion of the $F_2$ offspring were recessive homozygotes? How many different kinds of genotypes and phenotypes were generated in the $F_2$ offspring? What would your answer be if the geneticist testcrossed the decahybrid instead?

**27.** Consider the following crosses in pea plants and determine the genotypes of the parents in each cross. Yellow and green refer to seed color; tall and short refer to plant height.

| | **Progeny** | | | |
|---|---|---|---|---|
| **Cross** | **Yellow, Tall** | **Yellow, Short** | **Green, Tall** | **Green, Short** |
| a. Yellow, tall × yellow, tall | 89 | 31 | 33 | 10 |
| b. Yellow, short × yellow, short | 0 | 42 | 0 | 15 |
| c. Green, tall × yellow, short | 21 | 20 | 24 | 22 |

**28.** A brown-eyed, long-winged fly is mated with a red-eyed, long-winged fly. The progeny are

51 long, red 18 short, red
53 long, brown 16 short, brown

What are the genotypes of the parents?

**29.** True-breeding flies with long wings and dark bodies are mated with true-breeding flies with short wings and tan bodies. All the $F_1$ progeny have long wings and tan bodies. The $F_1$ progeny are allowed to mate and produce:

44 tan, long 14 tan, short
16 dark, long 6 dark, short

What is the mode of inheritance?

**30.** In peas, tall (*T*) is dominant to short (*t*), yellow (*Y*) is dominant to green (*y*), and round (*R*) is dominant to wrinkled (*r*). From a cross of two triple heterozygotes, what is the chance of getting a plant that is

**a.** tall, yellow, round?
**b.** short, green, wrinkled?
**c.** short, green, round?

**31.** In corn, the genotype *A- C- R-* is colored. Individuals homozygous for at least one recessive allele are colorless. Consider the following crosses involving colored plants, all with the same genotype. Based on the results, deduce the genotypes of the colored plants.

colored × *aa cc RR* → 1/2 colored; 1/2 colorless
colored × *aa CC rr* → 1/4 colored; 3/4 colorless
colored × *AA cc rr* → 1/2 colored; 1/2 colorless

**32.** Consider the following crosses in *Drosophila.* Based on the results, deduce which alleles are dominant and the genotypes of the parents. Orange and red are eye colors; crossveins occur on the wings.

| | **Progeny** | | | |
|---|---|---|---|---|
| **Parents** | **Orange, crossveins** | **Orange, crossveinless** | **Red, crossveins** | **Red crossveinless** |
| a. Orange, crossveins × orange, crossveins | 83 | 26 | 0 | 0 |
| b. Red, crossveins × red, crossveinless | 20 | 18 | 65 | 63 |
| c. Red, crossveinless × red, crossveins | 0 | 0 | 74 | 81 |
| d. Red, crossveins × red, crossveins | 28 | 11 | 93 | 34 |

**33.** In *Drosophila melanogaster,* a recessive autosomal gene, *ebony,* produces a dark body color when homozygous, and an independently assorting autosomal gene, *black,* has a similar effect. If homozygous *ebony* flies are crossed with homozygous *black* flies,

**a.** what will be the phenotype of the $F_1$ flies?

**b.** what phenotypes and what proportions would occur in the $F_2$ generation?

**c.** what phenotypic ratios would you expect to find in the progeny of the backcrosses of $F_1 \times$ *ebony*? $F_1 \times$ *black*?

**34.** *A, B,* and *C* are independently assorting Mendelian factors (genes) controlling the production of black pigment in a rodent species. Alleles of these genes are indicated as *a, b,* and *c,* respectively. Assume that *A, B,* and *C* act in this pathway:

$$\begin{tikzcd}
\text{colorless} \arrow[r, "\text{control by } A"] & \text{red} \arrow[d] & \\
& + \arrow[r, "\text{control by } C"] & \text{black} \\
\text{colorless} \arrow[r, "\text{control by } B"] & \text{red} \arrow[u] &
\end{tikzcd}$$

A black *AA BB CC* individual is crossed with a colorless *aa bb cc* to give black $F_1$ individuals. The $F_1$ individuals are selfed to give $F_2$ progeny.

**a.** What proportion of the $F_2$ generation is colorless?

**b.** What proportion of the $F_2$ generation is red?

**35.** In a particular *Drosophila* species, there are four strains differing in eye color: wild-type, orange-1, orange-2, and pink. The following matings of true-breeding individuals were performed.

| Cross | $F_1$ |
|---|---|
| wild-type × orange-1 | all wild-type |
| wild-type × orange-2 | all wild-type |
| orange-1 × orange-2 | all wild-type |
| orange-2 × pink | all orange-2 |
| $F_1$ (orange-1 × orange-2) × pink | 1/4 orange-2:<br>1/4 pink:1/4 orange-1:<br>1/4 wild-type |

What $F_2$ ratio would you expect if the $F_1$ progeny from orange-1 × orange-2 were selfed?

### GENOTYPIC INTERACTIONS

**36.** In a variety of onions, three bulb colors segregate: red, yellow, and white. A plant with a red bulb is crossed to a plant with a white bulb, and all the offspring have red bulbs. When these are selfed, the following plants are obtained:

| | |
|---|---|
| Red-bulbed | 119 |
| Yellow-bulbed | 32 |
| White-bulbed | 9 |

What is the mode of inheritance of bulb color, and how do you account for the ratio?

**37.** When studying an inherited phenomenon, a geneticist discovers a phenotypic ratio of 9:6:1 among offspring of a given mating. Give a simple, genetic explanation for this result. How would you test this hypothesis?

**38.** You notice a rooster with a pea comb and a hen with a rose comb in your chicken coop. Outline how you would determine the nature of the genetic control of comb type. How would you proceed if both your rooster and hen had rose combs?

**39.** Suggest possible mechanisms for the epistatic ratios given in table 2.4. Can you add any further ratios?

**40.** What are the differences among dominance, epistasis, and pleiotropy? How can you determine that pleiotropic effects, such as those seen in sickle-cell anemia, are not due to different genes?

**41.** You are working with the exotic organism *Phobia laboris* and are interested in obtaining mutants that work hard. Normal phobes are lazy. Perseverance finally pays off, and you successfully isolate a true-breeding line of hard workers. You begin a detailed genetic analysis of this trait. To date you have obtained the following results:

hard worker × nonworker
↓
$F_1$: all nonworkers of both sexes
$F_1$ female × worker male
↓
3/4 hard workers:1/4 nonworkers of both sexes

From these results, predict the expected phenotypic ratio from crossing two $F_1$ nonworkers.

### BIOCHEMICAL GENETICS

**42.** The following is a pathway from substance Q to substance U, with each step numbered:

$$\mathrm{Q} \xrightarrow{1} \mathrm{R} \xrightarrow{2} \mathrm{S} \xrightarrow{3} \mathrm{T} \xrightarrow{4} \mathrm{U}$$

Which product should build up in the cell and which products should never appear if the pathway is blocked at point 1? At 2? At 3? At 4?

**43.** The following chart shows the growth (+) or lack of growth (−) of four mutant strains of *Neurospora* with various additives. The additives are in the pathway of niacin biosynthesis. Diagram the pathway and show which steps the various mutants block. Which compound would each mutant accumulate? When you complete this problem, compare your results with figure 2.27. What effect on growth would you observe following a mutation in the pathway of serine biosynthesis?

| | **Mutants** | | | |
|---|---|---|---|---|
| **Additives** | **1** | **2** | **3** | **4** |
| Nothing | − | − | − | − |
| Niacin | + | + | + | + |
| Tryptophan | + | + | − | − |
| Kynurenine | + | + | − | − |
| 3-Hydroxyanthranilic acid | + | + | + | − |
| Indole | − | + | − | − |

**44.** The following shows the growth (+) or lack of growth (−) of various mutants in another biosynthesis pathway. Determine this pathway, the point of blockage for each mutant, and the substrate each mutant accumulates.

| | **Mutants** | | | | |
|---|---|---|---|---|---|
| **Additives** | **1** | **2** | **3** | **4** | **5** |
| Nothing | − | − | − | − | − |
| A | − | + | + | + | + |
| B | − | + | − | + | − |
| C | − | + | − | + | + |
| D | − | − | − | + | − |
| E | + | + | + | + | + |

**45.** Maple sugar urine disease is a rare inborn error of human metabolism in which the urine of affected individuals smells like maple sugar.

**a.** If two unaffected individuals have an affected child, what is the probable mode of inheritance of the disease?

**b.** What is the chance that the second child will be unaffected?

## CRITICAL THINKING QUESTIONS

**1.** In the shepherd's purse plant, the seed capsule comes in two forms, triangular and rounded. If two dihybrids are crossed, the resulting ratio of capsules is 15:1 in favor of triangular seed capsules. What type of biochemical pathway might generate that ratio?

**2.** Assume Mendel made the cross of two true-breeding plants that differed in all seven traits under study, one with all dominant traits, the other with all recessive traits. What would the ratio of phenotypes be in the $F_2$ generation?

*Suggested Readings for chapter 2 are on page B-1.*

# 3

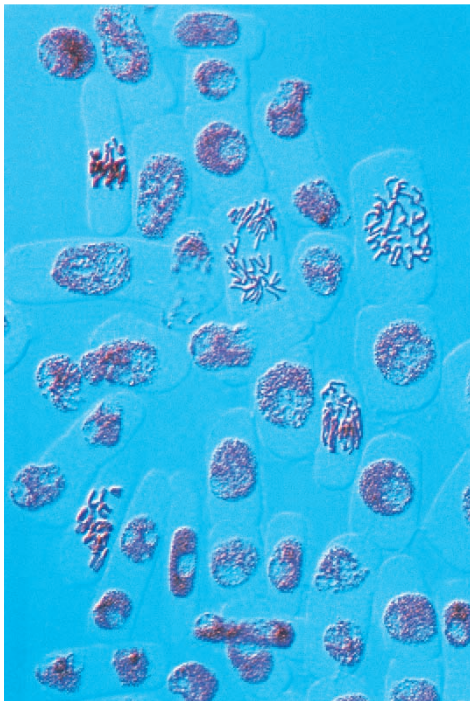
# MITOSIS AND MEIOSIS

## STUDY OBJECTIVES

1. To observe the morphology of chromosomes 48
2. To understand the processes of mitosis and meiosis 50
3. To analyze the relationships between meiosis and Mendel's rules 61

## STUDY OUTLINE

**Chromosomes** 48
**The Cell Cycle** 50
**Mitosis** 52
  The Mitotic Spindle 52
  Prophase 53
  Metaphase 54
  Anaphase 54
  Telophase 54
  The Significance of Mitosis 55
**Meiosis** 55
  Prophase I 56
  Metaphase I and Anaphase I 59
  Telophase I and Prophase II 59
  Meiosis II 60
  The Significance of Meiosis 61
**Meiosis in Animals** 63
**Life Cycles** 64
**Chromosomal Theory of Heredity** 66
*Summary* 66
*Solved Problems* 67
*Exercises and Problems* 67
*Critical Thinking Questions* 69



Onion (*Allium cepa*) cells in various stages of mitosis.

(© Andrew Syred/Tony Stone Images.)

The zygote, or fertilized egg of higher organisms, is the starting point of most life cycles. This zygote then divides many times to produce an adult organism. In animals, the adults then produce gametes that combine to start the cycle again. In higher plants, the adult is a sporophyte that produces spores by genetic reduction. These spores develop into gametophytes, which may or may not be independent, and gametophytes produce gametes that fuse to form the zygote (fig. 3.1). (Numerous variations on these themes exist, some of which we will discuss later in this chapter or others.) The process of cell division includes a nuclear and a cytoplasmic component. Nuclear division **(karyokinesis)** has two forms, a nonreductional **mitosis** in which the mother and daughter cells have exactly the same genetic complement, and a reductional **meiosis** in which the products, gametes in animals and spores in higher plants, have approximately half the genetic material as the parent cell. Halving the amount ensures that, when the gametes recombine, the amount of genetic material in a zygote is the same from generation to generation. The division of the cytoplasm, resulting in two cells from one original cell, is termed **cytokinesis.** In this chapter, we examine the processes of mitosis and meiosis, which allow chromosomes, the gene vehicles, to properly apportion among daughter cells. We will discuss the engineering difficulties these processes pose and the relationship of meiosis to Mendel's rules.

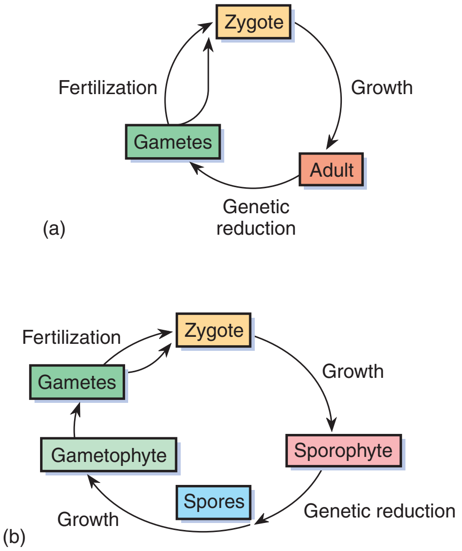


**Figure 3.1** Generalized life cycle of (*a*) animals and (*b*) plants.

Mendel's work was rediscovered at the turn of the century after being ignored for thirty-four years. One of the major reasons scientists could appreciate it in 1900 was that many of the processes that chromosomes undergo had been described. With those discoveries, a physical basis for genes had been found. That is, chromosomal behavior during gamete formation precisely fits Mendel's predictions for gene behavior during gamete formation. In this chapter, we look at the morphology of chromosomes and their behavior during somatic-cell division and gamete and spore formation.

Modern biologists classify organisms into two major categories: **eukaryotes,** organisms that have true, membrane-bound nuclei, and **prokaryotes,** organisms that lack true nuclei (table 3.1). Bacteria and blue-green algae are prokaryotes. All other organisms are eukaryotes. In most prokaryotes, the genetic material is a circle of double-stranded DNA (deoxyribonucleic acid) with some associated proteins; ancillary circles of double-stranded DNA called plasmids are also found frequently (see chapters 13 and 17). In eukaryotes, the genetic material, located in the nucleus (fig. 3.2), is linear, double-stranded DNA highly complexed with protein **(nucleoprotein).** In this chapter, we concentrate on the nuclear division processes of eukaryotes.

**Table 3.1** **Differences Between Prokaryotic and Eukaryotic Cells**

| | Prokaryotic Cells | Eukaryotic Cells |
|---|---|---|
| **Taxonomic groups** | Bacteria | All plants, fungi, animals, protists |
| **Size*** | Usually less than 5 μm in greatest dimension | Usually greater than 5 μm in smallest dimension |
| **Nucleus** | No true nucleus, no nuclear membrane | Nuclear membrane |
| **Genetic material** | One circular molecule of DNA, little protein | Linear DNA molecules complexed with histones |
| **Mitosis and meiosis** | Absent | Present |

* See table 3.2 on page 48.

**Table 3.2** **Metric Units of Linear Measurement**

| Unit | Abbreviation | Size |
|---|---|---|
| meter | m | 39.37 U.S. inches |
| centimeter | cm | $10^{-2}$ meter |
| millimeter | mm | $10^{-3}$ meter |
| micrometer | μm | $10^{-6}$ meter |
| nanometer | nm | $10^{-9}$ meter |
| Angstrom | Å | $10^{-10}$ meter |

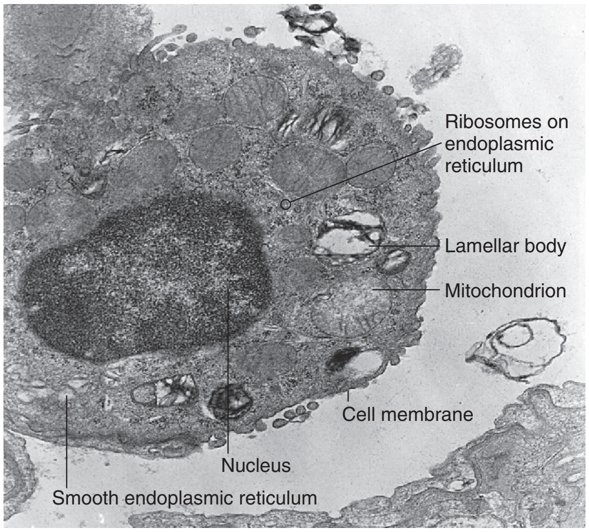

Ribosomes on endoplasmic reticulum
Lamellar body
Mitochondrion
Cell membrane
Nucleus
Smooth endoplasmic reticulum

**Figure 3.2** Mouse lung cell magnified 4,270x. (Courtesy of Wayne Rosenkrans.)

## CHROMOSOMES

Chromosomes were discovered by C. von Nägeli in 1842. The term **chromosome,** which W. Waldeyer coined in 1888, means "colored body." Von Nägeli discovered chromosomes after staining techniques were developed that made them visible. The nucleoprotein material of the chromosomes is referred to as **chromatin.** When diffuse, chromatin is referred to as **euchromatin;** when condensed and readily visible, as **heterochromatin.**

Although all eukaryotes have chromosomes, in the **interphase** between divisions, they are spread out or diffused throughout the nucleus and are usually not identifiable. Each chromosome, with very few exceptions, has a distinct attachment point for fibers **(microtubules)** that make up the mitotic and meiotic spindle apparatuses. The attachment point occurs at a constriction in the chromosome termed the **centromere,** which is composed of several specific DNA sequences (see chapter 15). The **kinetochore** is the proteinaceous structure on the surface of the centromere to which microtubules of the spindle attach. Chromosomes can be classified according to whether the centromere is in the middle of the chromosome **(metacentric),** at the end of the chromosome **(telocentric),** very near the end of the chromosome **(acrocentric),** or somewhere in between (**subtelocentric** or **submetacentric;** figs. 3.3 and 3.4). For any particular chromosome, the position of the centromere is fixed. In various types of preparations, dark bands **(chromomeres)** are visible (see chapter 15).

Most higher eukaryotic cells are **diploid;** that is, all their chromosomes occur in pairs. One member of each pair came from each parent. **Haploid** cells, which include the reproductive cells (gametes), have only one copy of each chromosome. In the diploid state, members of the same chromosome pair are referred to as **homologous chromosomes** (homologues); the two make up a homologous pair.

The total chromosomal complement of a cell, the **karyotype,** can be photographed during mitosis and rearranged in pairs to make a picture called a karyotype or **idiogram** (fig. 3.5). From the idiogram it is possible to see whether the chromosomes have any abnormalities and to identify the sex of the organism. As you can see from figure 3.5, all of the homologous pairs are made up of identical partners, and are thus referred to as **homomorphic chromosome pairs.** A potential exception is the sex chromosomes, which in some species are of un-

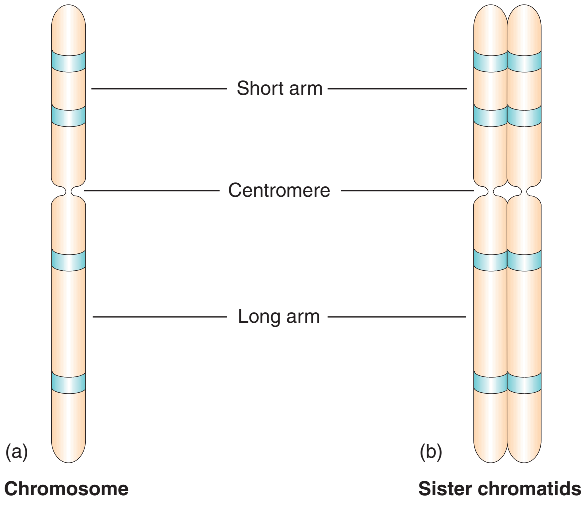


**Figure 3.3** (*a*) Submetacentric chromosome and (*b*) submetacentric chromosome in mitosis. The chromosome is best seen after it has duplicated but before the identical halves (sister chromatids) separate.

equal size and are therefore called a **heteromorphic chromosome pair.**

The number of chromosomes individuals of a particular species possess is constant. Some species exist mostly in the haploid state or have long haploid intervals in their life cycle. For example, pink bread mold, *Neurospora crassa,* a fungus, has a chromosome number of seven ($n = 7$) in the haploid state. Its diploid number is, of course, fourteen ($2n = 14$). The diploid chromosome numbers of several species appear in table 3.3.

In eukaryotes, two processes partition the genetic material into offspring, or daughter, cells. One is the simple division of one cell into two. In this process, the two daughter cells must each receive an exact copy of the genetic material in the parent cell. The cellular process is simple cell division, and the nuclear process accompanying it is mitosis. In the other partitioning process, the genetic material must precisely halve so that fertilization will restore the diploid complement. The cellular process is gamete formation in animals and spore formation in higher plants, and the nuclear process is meiosis. The term *mitosis* comes from the Greek word for "thread," referring to a chromosome. The term *meiosis* comes from the Greek meaning "to lessen."

Chromosomes separate in both processes of nuclear division. The division of the cytoplasm of the cell, *cytokinesis,*

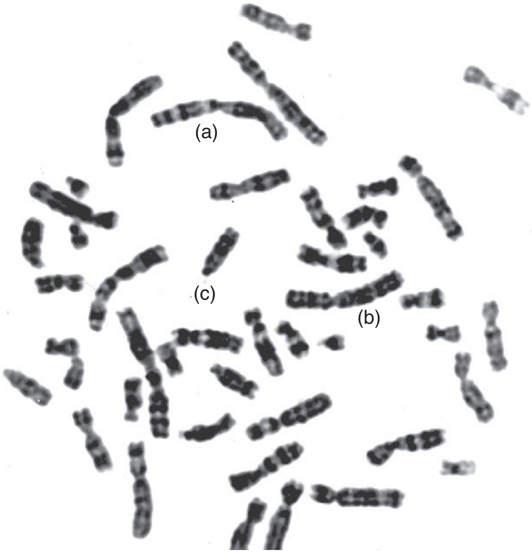


**Figure 3.4** (*a*) Metacentric, (*b*) submetacentric, and (*c*) acrocentric chromosomes in human beings. Except in telocentric chromosomes, the centromere divides the chromosome into two arms. (Reproduced courtesy of Dr. Thomas G. Brewster, Foundation for Blood Research, Scarborough, Maine.)

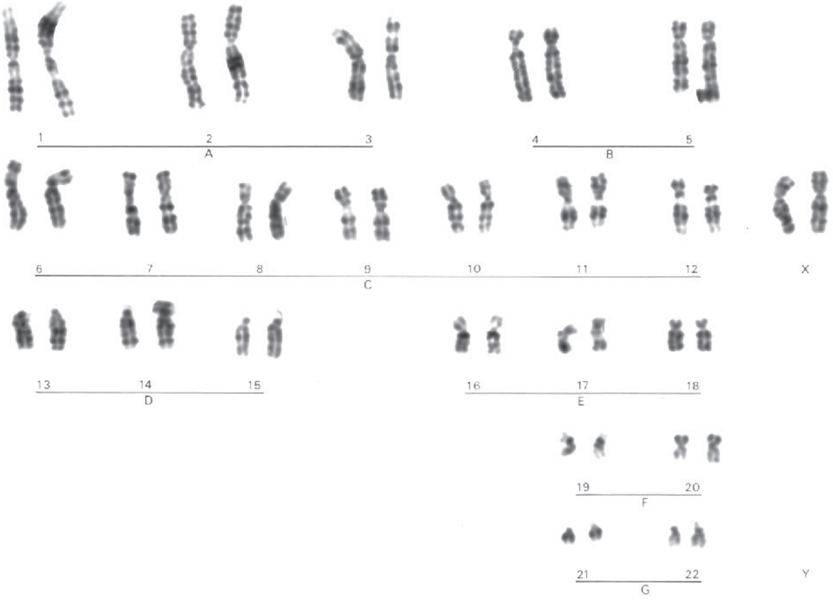


**Figure 3.5** Idiogram or karyotype of a human female (two X chromosomes, no Y chromosome). A male would have one X and one Y chromosome. The chromosomes are grouped into categories (A–G, X, Y) by length and centromere position. Similar chromosomes are often distinguished by their chromomeres. (Reproduced courtesy of Dr. Thomas G. Brewster, Foundation for Blood Research, Scarborough, Maine.)

**Table 3.3** **Chromosome Number for Selected Species**

| Species | 2*n* |
|---|---|
| Human being (*Homo sapiens*) | 46 |
| Garden pea (*Pisum sativum*) | 14 |
| Fruit fly (*Drosophila melanogaster*) | 8 |
| House mouse (*Mus musculus*) | 40 |
| Roundworm (*Ascaris* sp.) | 2 |
| Pigeon (*Columba livia*) | 80 |
| Boa constrictor (*Constrictor constrictor*) | 36 |
| Cricket (*Gryllus domesticus*) | 22 |
| Lily (*Lilium longiflorum*) | 24 |
| Indian fern (*Ophioglossum reticulatum*) | 1,260 |

*Note:* 2*n* is the diploid complement. The fern has the highest known diploid chromosome number.

is less organized. In animals, a constriction of the cell membrane distributes the cytoplasm. In plants, the growth of a cell plate accomplishes the same purpose.

## THE CELL CYCLE

The continuity of life depends on cells growing, replicating their genetic material, and then dividing, a process called the **cell cycle** (fig. 3.6). Although cells usually divide when they have doubled in volume, the control of this process is very complex and precise. Not only do all the steps have to occur in sequence, but the cell must also "know" when to proceed and when to wait. Continuing at inappropriate moments—for example, before the DNA has replicated or when the chromosomes or spindle are damaged—could have catastrophic consequences to a cell or a whole organism. Numerous stops occur during the cycle to assess whether the next step should proceed.

Early research into the cell cycle involved fusing cells in different stages of the cycle (such as the $G_1$, S, and $G_2$ phases; see fig. 3.6) to determine whether the cytoplasmic components of one cell would affect the behavior of the other. Results of these experiments led to the discovery of a protein complex called the **maturation-promoting factor (MPF)** because of its role in causing oocytes to mature. It is now also referred to as the **mitosis-promoting factor** since it initiates the mitosis phase of the cell cycle. Further research has shown that MPF is made of two proteins, one that oscillates in quantity during the cell cycle and one whose quantity is constant. The oscillating component is referred to as **cyclin;** the constant gene product is an enzyme controlled by the *cdc2* gene (*cdc* stands for *c*ell *d*ivision *c*ycle) called Cdc2p. Cdc2p is a kinase, an enzyme that phosphorylates other proteins, transferring a phosphate group from ATP to an amino acid of the protein it is acting on. (Phosphorylation controls many of the processes in mitosis and in metabolism in general; for example, the nuclear membrane begins to break down when its subunits are phosphorylated.) Because the Cdc2p kinase works when combined with cyclin, it is referred to as a **cyclin-dependent kinase (CDK).** Several of these kinase-cyclin combinations control stages of the cell cycle; the cyclin of the mitosis-promoting factor is called cyclin B. In general, cylin-dependent kinases are regulated by phosphorylation and dephosphorylation, cyclin levels, and activation or deactivation of inhibitors.

Normally, Cdc2p remains at high levels in the cell but does not initiate mitosis for two reasons. First, phosphate groups block its active site, the place on the enzyme that actually does the phosphorylating. Second, the enzyme can only function when it combines with a molecule of cyclin B, the protein that oscillates during the cell cycle. Cyclin B is at very low levels when mitosis ends. During ensuing cell growth, numbers of cyclin B molecules increase, combining with Cdc2p proteins until a critical quantity is reached. However, Cdc2p-cyclin B complexes are still not active. That requires the product of another gene to dephosphorylate the Cdc2p-cyclin B complex. At that point, the Cdc2p-cyclin B complex goes into action, initiating the changes that begin mitosis (fig. 3.7). Presumably the cell is ready for mitosis at that point, having

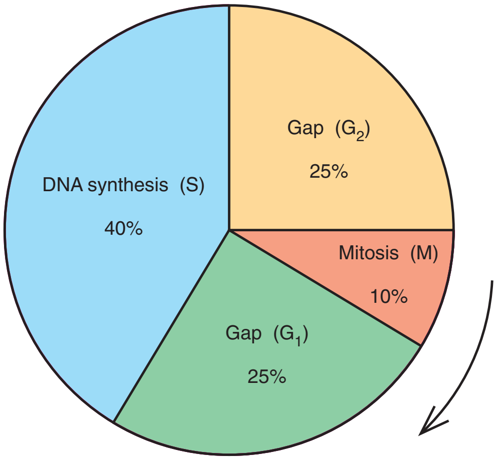


**Figure 3.6** Cell cycle in the broad bean, *Vicia faba*. Total time in the cycle is under twenty hours. The DNA content of the cell doubles during the S phase and is then reduced back to its original value by mitosis.

gone through G1, S, and G2 phases (which we will discuss in detail later in the chapter).

Once mitosis has been initiated, cyclin B, along with other proteins that have served their purpose by this point in the cell cycle, breaks down with the help of a protein complex called the **anaphase-promoting complex (APC),** also called the **cyclosome.** The cyclosome works by attaching a **ubiquitin** molecule to the proteins that are to be broken down. (Ubiquitin is a polypeptide of 76 amino acids; it directs the attached protein into a breakdown pathway discussed in chapter 16.) Cdc2p is then phosphorylated to block its active site. The cell now completes mitosis and enters $G_1$; quantities of cyclin B are very low, and virtually no functioning Cdc2p-cyclin B remains (fig. 3.7). Thus, active Cdc2p is the kinase that controls the initiation of mitosis.

Some points in the cell cycle, such as the initiation of mitosis, can be delayed until all necessary conditions are in place. These **checkpoints** allow the cell to make sure that various events have been "checked off" as completed before the next phase begins. **Surveillance mechanisms** that involve dozens of proteins, many just discovered, oversee these checkpoints. In the cell cycle, three checkpoints involve cyclin-dependent kinases; each has its own specific cyclin that initiates either the $G_1$, S, or mitosis phase. In addition, other checkpoints that don't involve cyclin-dependent kinases occur at other transition phases in the cell cycle.

Cell cycle control is of particular interest because the cell cycle routinely halts if there is genetic damage, giving the cell a chance to repair the damage before committing to cell division. If the damage is too extreme, the cell can enter a programmed cell death sequence, discussed in chapter 16. If these mechanisms fail, cancer may result. The genetic control of the cell cycle is one of the most active areas of current research.

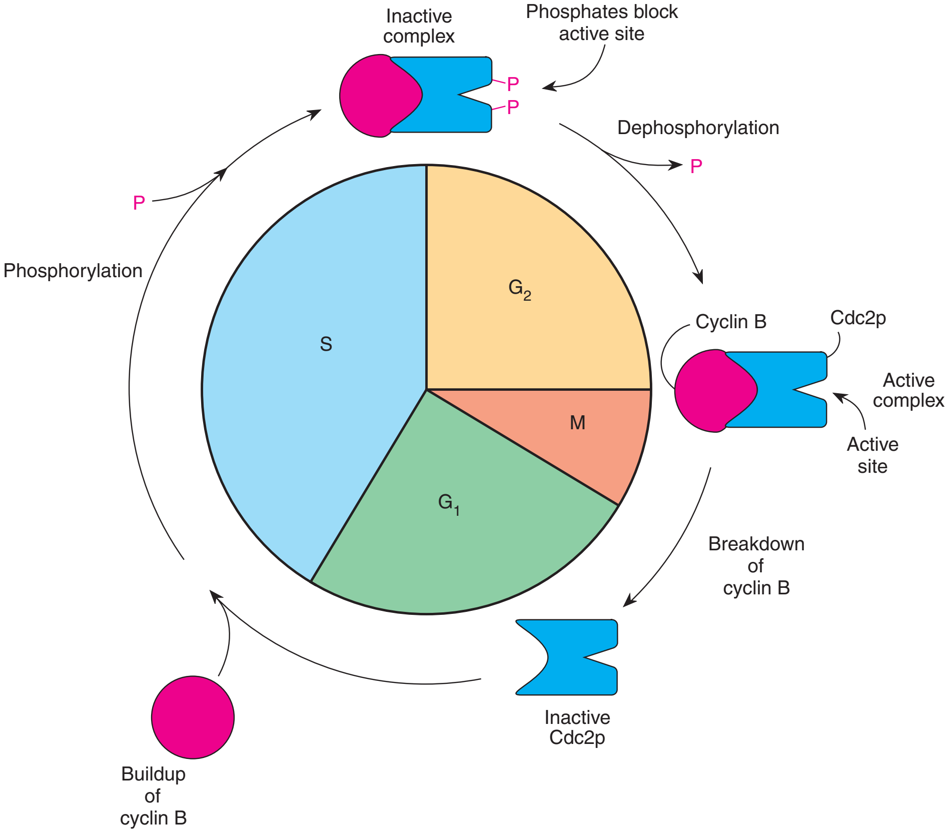


**Figure 3.7** The proteins Cdc2p (CDK1) and cyclin B combine to form the maturation-promoting (or mitosis-promoting) factor. During mitosis, cyclin B is broken down. During $G_1$ and S phases, cyclin B builds up and combines with Cdc2P, which is then phosphorylated at the active site to render it inactive. Dephosphorylation, a process that begins to take place only after DNA replication is finished, produces an active maturation-promoting factor.

# MITOSIS

Consider the engineering problem that mitosis must solve. Identical **chromatids,** called **sister chromatids,** the result of chromosomal replication, must separate so that each goes into a different daughter cell (see fig. 3.3). These chromatids are the visible manifestation of the chromosomal replication that has taken place in the S phase of the cell cycle. The chromatids are initially held together; each will be called a chromosome when it separates and becomes independent. Each of the two daughter cells then ends up with a chromosome complement identical to that of the parent cell. Mitosis is nature's elegant process to achieve that end—surely an engineering marvel.

Mitosis is a continuous process. However, for descriptive purposes, we can break it into four stages: **prophase, metaphase, anaphase,** and **telophase** (Greek: *pro-,* before; *meta-,* mid; *ana-,* back; *telo-,* end). Replication (duplication) of the genetic material occurs during the S phase of the cell cycle (see fig. 3.6). The timing of the four stages varies from species to species, from organ to organ within a species, and even from cell to cell within a given cell type.

## The Mitotic Spindle

The process of mitosis involves an apparatus called the **spindle.** This structure is composed of microtubules, hollow cylinders made of protein subunits; each subunit is composed of one molecule of α tubulin and one of β tubulin; and each tubulin is the product of a different gene. (The spindle is named for the rounded rods, tapered at each end, once commonly used to hold yarn or thread.) Microtubules provide shape and structure to a eukaryotic cell as well as allow the cell to move its internal components and to move the cell itself with cilia and flagella. Motion occurs as the microtubules slide past each other, a vesicle of some kind slides along the microtubules, and the microtubules shorten. Two proteins make up the microtubule motors that allow motion: **kinesin** and **dynein.** Scientists have studied microtubules through protein chemistry, through mutant organisms, and through innovative methods such as by coupling tubulin subunits with fluorescing dyes to observe the microtubules in action.

Microtubules are in a dynamic equilibrium, with subunits constantly being added or removed at both ends. On any microtubule, more activity occurs at one end than the other. The more active end of the tubule is called the plus end, the less active end the minus end (fig. 3.8). Both ends may be adding or removing subunits, or the plus end may be adding while the minus end is removing subunits. Generally, dynein causes movement toward the minus end, whereas kinesin causes movement toward the plus end of a microtubule, although exceptions exist.

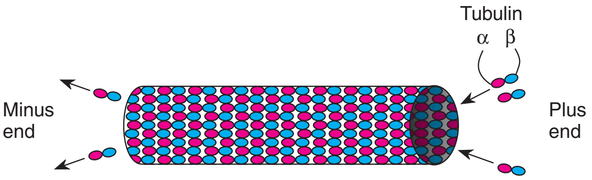


**Figure 3.8** Microtubules are hollow tubes made of α and β tubulin subunits that are constantly being added or removed.

Microtubules are formed from active centers called **microtubule organizing centers. Centrioles,** composed of two cylinders—themselves composed of microtubules—are microtubule organizing centers for cilia and flagella. Under those circumstances, the centrioles are referred to as **basal bodies.** The centrioles were also originally believed to organize spindles. However, for most organisms, the microtubule organizing center is called the **centrosome.** In some organisms, such as fungi, a different cell organelle, the **spindle pole body,** serves this function. In most animals, the centrosome contains a centriole (fig. 3.9). However, the centriole is absent in most higher plants. Moreover, innovative ex-

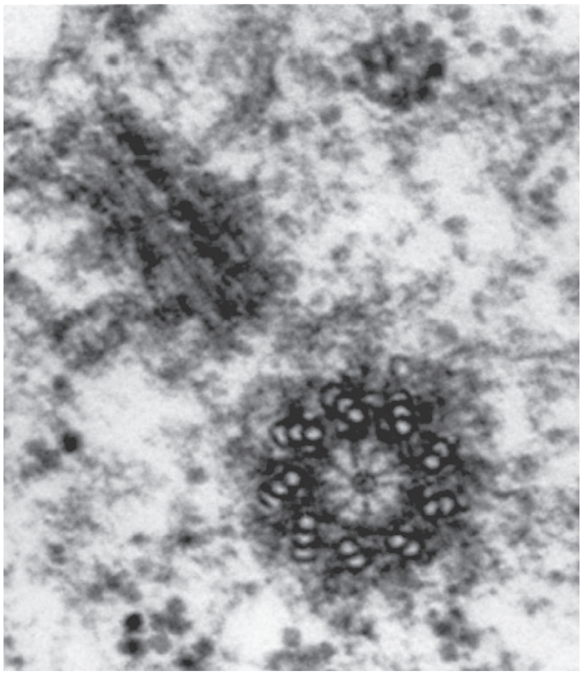


**Figure 3.9** A centriole is composed of two barrels at right angles to each other. Each barrel is composed of nine tripartite units and a central cartwheel. Each of the three parts of a tripartite unit is a microtubule. Magnification 111,800×. (Reproduced from *The Journal of Cell Biology,* 1968, Vol. 37, p. 381, by copyright permission of The Rockefeller University Press.)

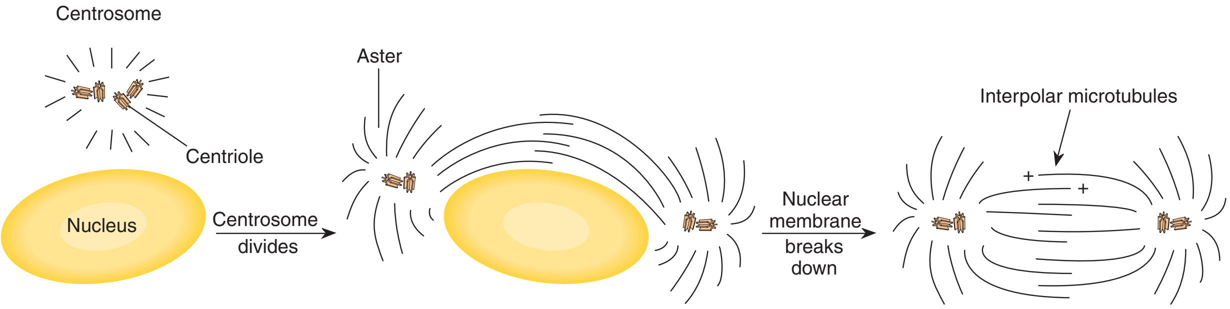


**Figure 3.10** Early in mitosis, the centrosome divides, and the separating halves move to opposite poles of the cell. This creates a spindle in the middle of the cell after the nuclear membrane breaks down.

periments that removed the centrioles from cells that normally had them demonstrated that the centriole is not necessary for spindle formation. So, although we used to believe that the centriole formed the spindle in many organisms, we now know that the spindle is usually organized around the centrosome, which can function in this capacity without a centriole.

The centriole, when present, replicates during the S and $G_2$ phases. When mitosis begins, the centrosome divides and moves to opposite poles of the cell, around the nucleus (fig. 3.10). The centrosomes trail microtubules, forming the spindle, that at this point begin at each centrosome and overlap in the middle of the cell. These are called **interpolar microtubules.** Microtubules also spread out from the centrosome in the opposite direction from the spindle itself, forming an **aster** (see fig. 3.10). The minus ends of microtubules emanate from the centrosome and the plus ends overlap in the middle of the cell. A third form of tubulin, $\gamma$ tubulin, is needed to begin the formation of a microtubule.

## Prophase

This stage of mitosis is characterized by the formation of the spindle and a shortening and thickening of the chromosomes so that individual chromosomes become visible. (We will discuss details of the molecular structure of the eukaryotic chromosome and the processes of coiling and shortening in chapter 15.) At this time also, the nuclear envelope (membrane) disintegrates and the **nucleolus** disappears (fig. 3.11). The nucleolus is a darkly stained body in the nucleus that is involved in ribosome construction and that forms around a **nucleolar organizer** locus on one of the chromosome pairs. The number of nucleoli varies in different species, but in the simplest case there are two nucleolar organizers per nucleus, one each on the two members of a homologous pair of chromosomes. Nucleoli re-form after mitosis.

As prophase progresses, each chromosome is composed of two identical (sister) chromatids (see fig. 3.3); the chromosomes continue to shorten and thicken. The centromeres have already divided, and no new DNA synthesis is needed for the process to be completed. At this point, the sister chromatids are kept together by a complex, called **cohesin,** made up of at least four different protiens.

Spindle fibers are initially nucleated at the centrosome and grow outward into the cytoplasm (fig. 3.12). Some of these fibers "capture" a kinetochore, the proteinaceous complex at the centromere of each sister chromatid; these fibers are called **kinetochore microtubules.** At first, one kinetochore or the other randomly attaches to a spindle fiber. As the microtubules further move the chromosomes and as new microtubules attach and old microtubules break, each sister kinetochore eventually attaches to microtubules emanating from different poles. This ensures that sister chromatids move to opposite poles during anaphase. The number of microtubules that attach to each kinetochore differs in different species. It seems that 1 attaches to each kinetochore in yeast, 4 to 7 attach to each kinetochore in the cells of a rat fetus, and 70 to 150 attach in the plant *Haemanthus* (see fig. 3.12).

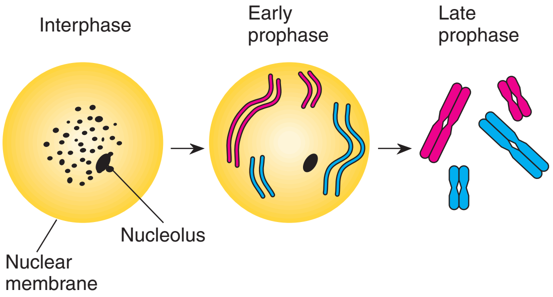


**Figure 3.11** Nuclear events during interphase and prophase of mitosis. In this cell, $2n = 4$, consisting of one pair of long and one pair of short metacentric chromosomes. Maternal chromosomes are *red;* paternal chromosomes are *blue.* Note that each chromosome consists of two chromatids when the cell enters mitosis.

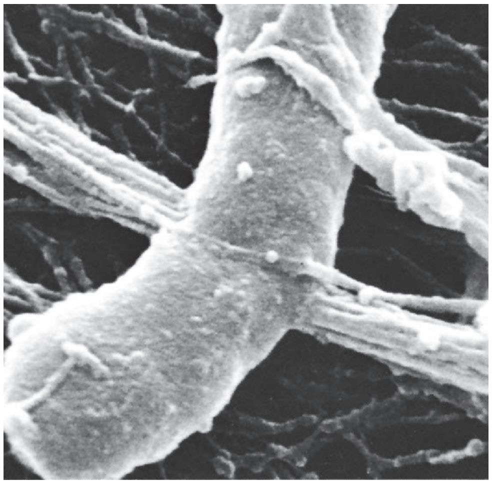


**Figure 3.12** Scanning electron micrograph of the centromeric region of a metaphase chromosome from the plant *Haemanthus katherinae*. Spindle fiber bundles on either side of the centromere extend in opposite directions. A fiber not connected to the kinetochore is visible lying over the centromere. These fibers are 60 to 70 nm in diameter. (Waheeb K. Heneen, "The centromeric region in the scanning electron microscope," *Hereditas,* 97 (1982): 311–14. Reproduced by permission.)

## Metaphase

During metaphase, the chromosomes move to the equator of the cell. With the attachment of the spindle fibers and the completion of the spindle itself, the chromosomes jockey into position in the equatorial plane of the spindle, called the **metaphase plate.** This happens as kinetochore microtubules exert opposing tension on the two sister kinetochores. Alignment of the chromosomes on this plate marks the end of metaphase (fig. 3.13).

## Anaphase

During anaphase, the sister chromatids separate and move toward opposite poles on the spindle. The physical separation of the sister chromatids and their movement to opposite poles are two separate activities. Chromatid separation represents a checkpoint in the process of mitosis; a surveillance mechanism will not allow the process to continue until all chromosomes are lined up on the metaphase plate with their sister kinetochores held by microtubules from opposite poles. The surveillance mechanism somehow checks the physical tension the spindle fibers exert on a pair of sister chromatids; an unpaired chromatid can delay or stop the process. Initially, an inhibitory protein called **securin** binds an enzyme called **separin** that can break down cohesin, the complex holding the chromatids together. At the correct moment, the cyclosome ubiquitinates the inhibitor, causing it to break down and freeing the separin to break down cohesin. This liberates the sister chromatids from each other (and is the instant when chromatids become chromosomes).

The spindle then separates the sister chromatids in two stages, called **anaphase A** and **anaphase B.** In anaphase A, the chromosomes move toward the poles (fig. 3.14). During this process, the kinetochore itself acts as a microtubule motor, disassembling microtubules as it moves down them, pulling the chromosomes along (fig. 3.15). Thus, metacentric chromosomes appear V-shaped (as in fig. 3.15), subtelocentrics appear J-shaped, and telocentrics appear rod-shaped. In anaphase B, the spindle itself elongates as overlapping interpolar microtubules slide apart. The general elongation of the spindle pulls the chromosomes apart.

## Telophase

At the end of anaphase (fig. 3.16), the separated sister chromatids (now full-fledged chromosomes) have been pulled to opposite poles of the cell. The cell now reverses the steps of prophase to return to the interphase state (fig. 3.17). The chromosomes uncoil and begin to direct protein synthesis. A nuclear envelope re-forms around each set of chromosomes, nucleoli re-form, and cytokinesis takes place. The spindle breaks down into tubulin subunits; a residual of microtubules remains at

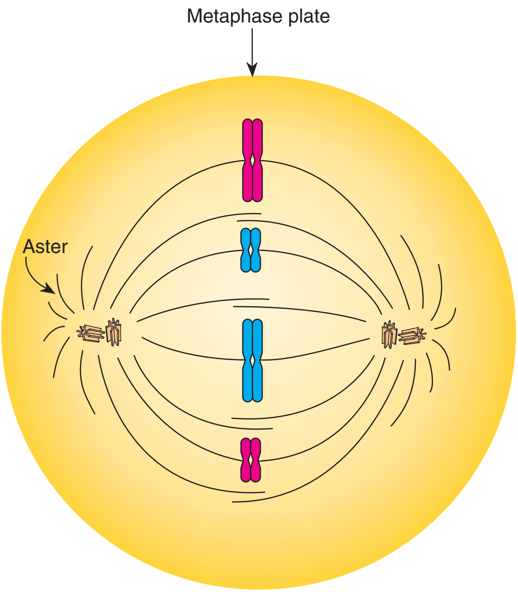


**Figure 3.13** Metaphase of mitosis. In this cell, $2n = 4$. Maternal chromosomes are *red;* paternal chromosomes are *blue*.

the center of the cell and seems to be involved in the formation of a constricting ring in animal cells or in the growth of a cell plate in plant cells. The cell has now entered the $G_1$ phase of the cell cycle (see fig. 3.6). Figure 3.18 summarizes mitosis.

### The Significance of Mitosis

Cytokinesis and mitosis result in two daughter cells, each with genetic material identical to that of the parent cell. This exact distribution of the genetic material, in the form of chromosomes, to the daughter cells, ensures the stability of cells and the inheritance of traits from one cell generation to the next. Cells have evolved complex life functions; with mitosis, they will produce offspring cells with these same capabilities. With this stability assured, single-celled organisms could thrive and multicellular organisms could evolve.

## MEIOSIS

Gamete formation presents an entirely new engineering problem to be solved. To form gametes in animals (and, for the most part, to form spores in plants), a diploid organism with two copies of each chromosome must form daughter cells that have only one copy of each chromosome. In other words, the genetic material must be reduced by half so that when gametes recombine to form zygotes, the original number of chromosomes is restored, not doubled.

If we were to try to engineer this task, we would first need to be able to recognize homologous chromosomes. We could then push one member of each pair into one daughter cell and the other into the other daughter cell. If we were unable to recognize homologues, we would not be able to ensure that each daughter cell received one and only one member of each pair. The cell solves this problem by pairing up homologous chromosomes during an extended prophase. The spindle apparatus then separates members of the homologous chromosome pairs, just as it separates sister chromatids during mitosis. But there is one complication. As in mitosis, cells entering meiosis have already replicated their chromosomes. Therefore, two nuclear divisions without an intervening chromosome replication are necessary to produce haploid gametes or

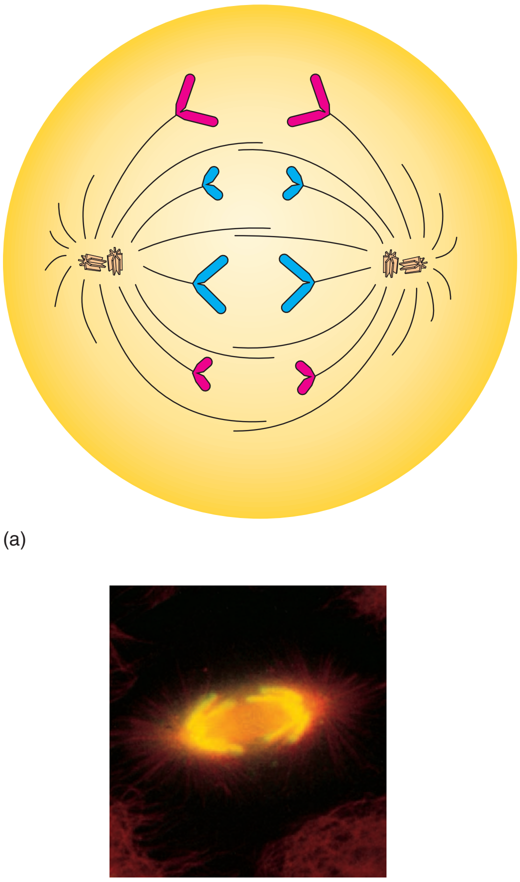


**Figure 3.14** (*a*) The mitotic spindle during anaphase. In this cell, $2n = 4$. Maternal chromosomes are *red;* paternal chromosomes are *blue*. (*b*) Fluorescent microscope image of a cultured cell in anaphase. Microtubles are *red;* chromosomes (DNA) are stained *yellow*. ([*b*] John M. Murray, Department of Anatomy, University of Pennsylvania. Cover of *BioTechniques,* volume 7, number 3, March 1989. Reproduced with permission.)

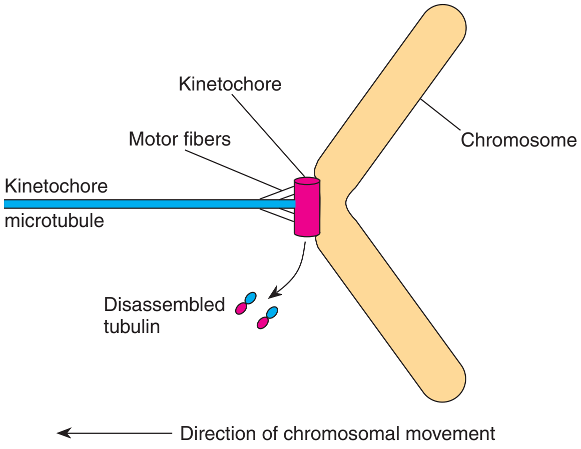


**Figure 3.15** The kinetochore acts as a microtubule motor, pulling the chromosome along the kinetochore microtubules toward the pole. One microtubule is shown, although many may be present.

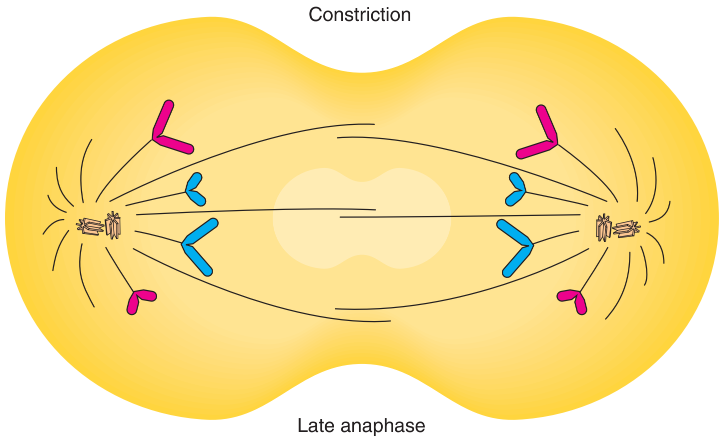


**Figure 3.16** Late anaphase of mitosis; $2n = 4$. A constriction begins to form in the middle of the cell (in animals). Maternal chromosomes are *red;* paternal chromosomes are *blue*.

spores. Meiosis is, then, a two-division process that produces four cells from each original parent cell. The two divisions are known as meiosis I and meiosis II.

Unlike mitosis, meiosis occurs only in certain kinds of cells. In animals, meiosis begins in the primary gametocytes; in higher plants, the process takes place only in the spore-mother cells of the sporophyte generation (see fig. 3.1). At the end of this chapter, we review the processes of gamete and spore formation in animals and plants, respectively.

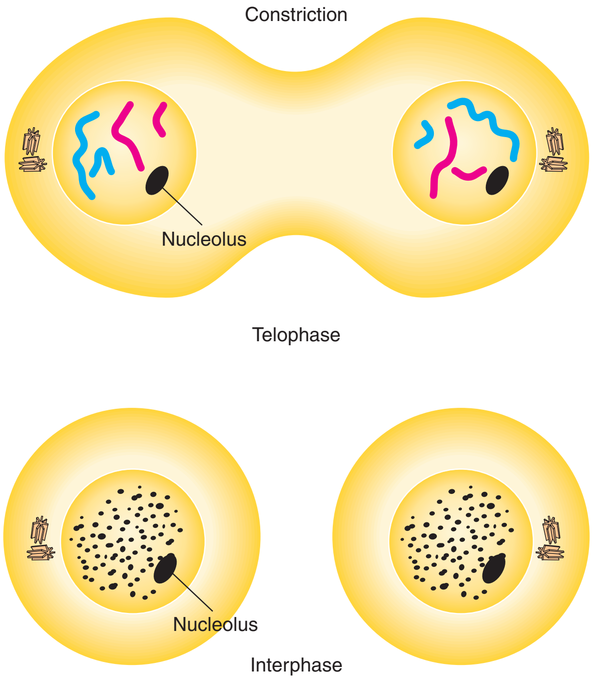


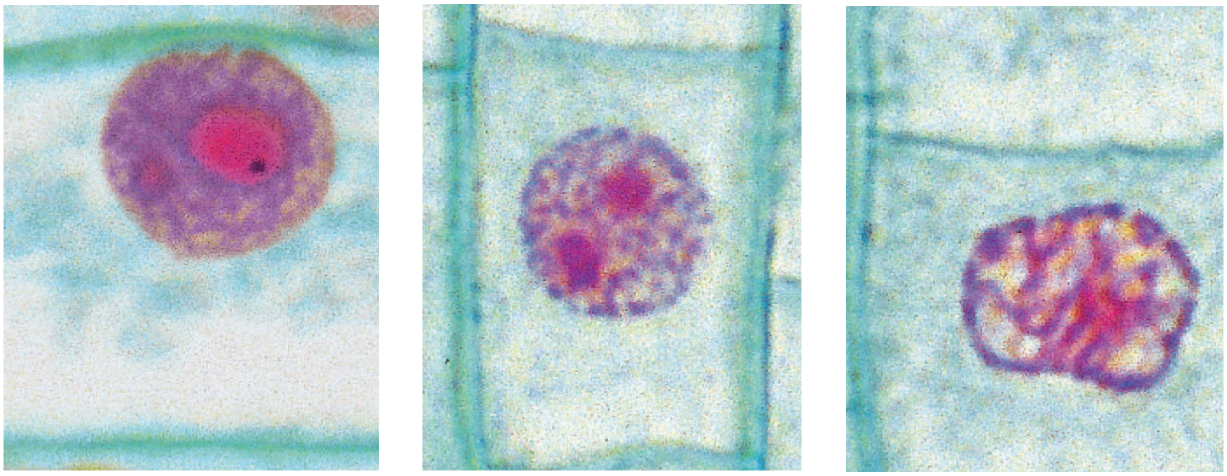
**Figure 3.17** Telophase and interphase of mitosis; $2n = 4$. Maternal chromosomes are *red;* paternal chromosomes are *blue*.

## *Prophase I*

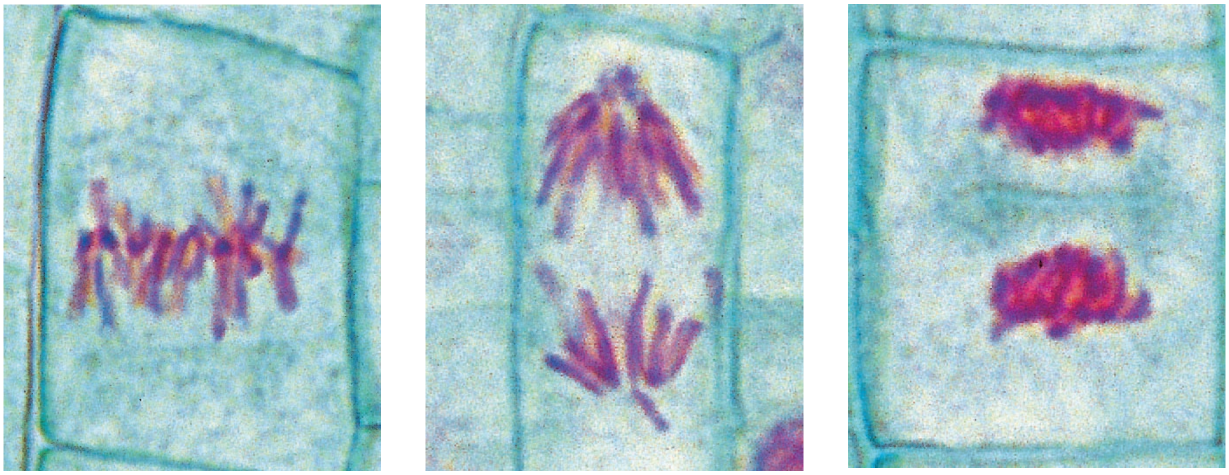
Cytogeneticists have divided the prophase of meiosis I into five stages: **leptonema, zygonema, pachynema, diplonema,** and **diakinesis** (Greek: *lepto-,* thin; *zygo-,* yoke-shaped; *pachy-,* thick; *diplo-,* double; *dia-,* across). A cell entering prophase I (leptotene stage) behaves similarly to one entering prophase of mitosis, with the centrosome duplicated and the spindle forming around the intact nucleus. (Note the adjectival forms—*leptotene*—versus the noun forms—*leptonema*—of the stage names.) As the chromosomes coil down in size during leptonema, they are visible as individual threads: sister chromatids are in such close apposition that they are not distinct. The chromosomes are more spread out than they are in mitosis, with dark spheres or bands called chromomeres interspersed.

The tips of the chromosomes are attached to the nuclear membrane in the leptotene stage (fig. 3.19). In the leptotene to zygotene transition, the tips of the chromosomes move until most end up in a limited region near each other. This forms an arrangement called a **bouquet stage.** Presumably, this arrangement helps homologous chromosomes find each other and begin the pairing process without becoming entangled.

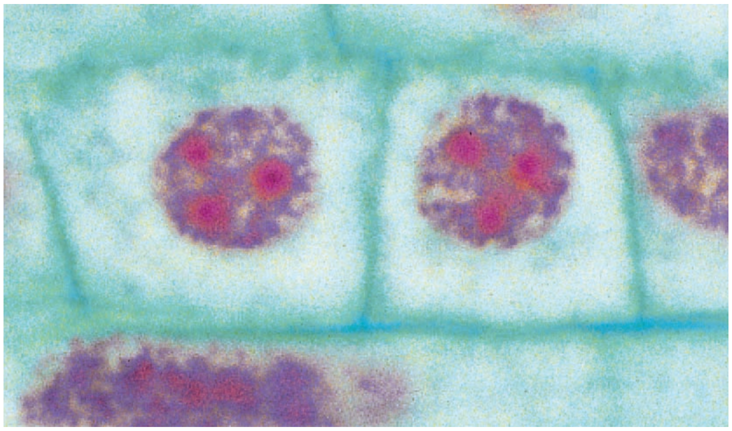
The pairing of homologous chromosomes marks the zygotene stage. Initial contact between identical regions of homologous chromosomes leads to a point-for-point pairing along their lengths. This process is referred to as **synapsis.** A proteinaceous complex, referred to as a



(a) Interphase (b) Early prophase (c) Late prophase



(d) Metaphase (e) Anaphase (f) Telophase



(g) Daughter cells

**Figure 3.18** Cells in interphase and in various stages of mitosis in the onion root tip. The average cell is about 50 $\mu$m long. (© The McGraw-Hill Companies, Inc./Kingsley Stern, photographer.)

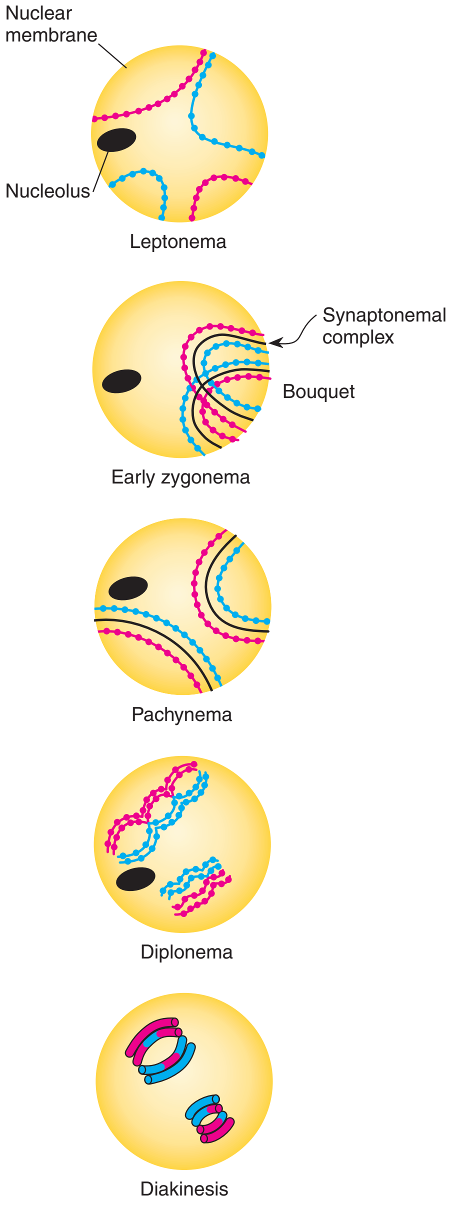

Nuclear membrane
Nucleolus
Leptonema
Synaptonemal complex
Bouquet
Early zygonema
Pachynema
Diplonema
Diakinesis

**Figure 3.19** Prophase I of meiosis; 2*n* = 4 (nuclei shown). Maternal chromosomes are *red;* paternal chromosomes are *blue*. Note that crossing over is evident at diplonema.

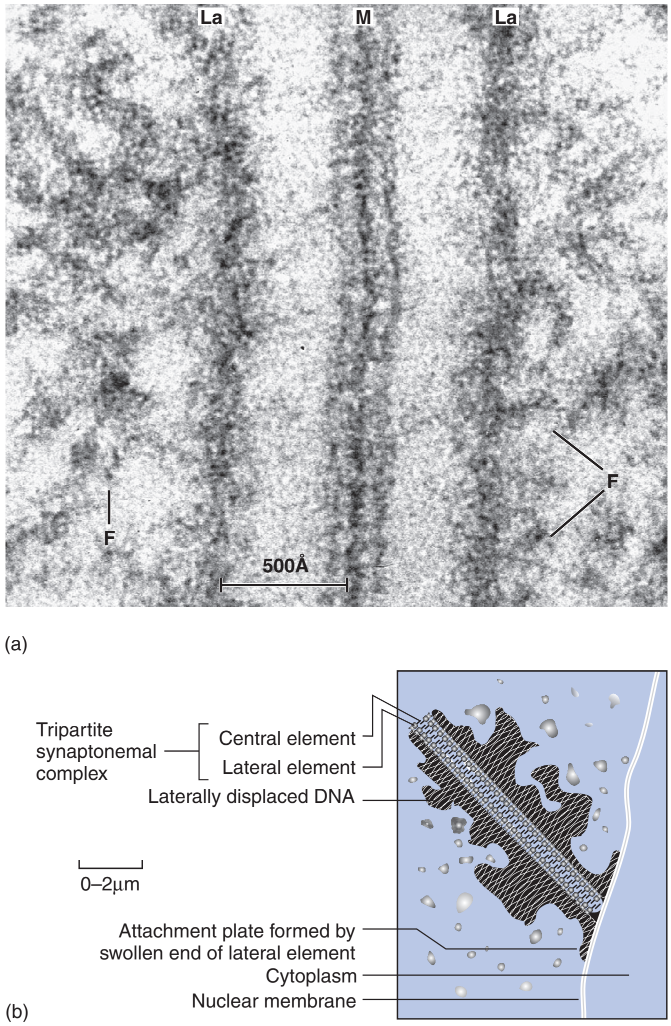

La
M
La
F
F
500Å
(a)
Tripartite synaptonemal complex
Central element
Lateral element
Laterally displaced DNA
0–2μm
Attachment plate formed by swollen end of lateral element
Cytoplasm
Nuclear membrane
(b)

**Figure 3.20** The synaptonemal complex. *(a)* In the electron micrograph, *M* is the central element, *La* are lateral elements, and *F* are chromosome fibers. Magnification 400,000×. *(b)* Diagram of the structure. ([*a*] R. Wettstein and J. R. Sotelo, "The molecular architecture of synaptonemal complexes," in E. J. DuPraw, ed., *Advances in Cell and Molecular Biology,* vol. 1 (New York: Academic Press, 1971), p. 118. Reproduced by permission. [*b*] From B. John and K. R. Lewis, *Chromosome Hierarchy.* Copyright © 1975 Oxford University Press, London, England. Reprinted by permission of the Oxford University Press.)

**synaptonemal complex** (fig. 3.20), appears between the homologous chromosomes and mediates synapsis in an unknown way. At this point, the chromosome figures are referred to as **bivalents,** one bivalent per homologous pair. The synapsis of all chromosomes marks the end of zygonema.

The chromosomes now continue to shorten and thicken, giving pachynema its name. During the entire prophase, **crossing over** takes place. When two chromatids come to lie in close proximity, enzymes can break both chromatid strands and reattach them differently (fig. 3.21). Thus, although genes have a fixed position on a chromosome, alleles that started out attached to a paternal centromere can end up attached to a maternal cen-

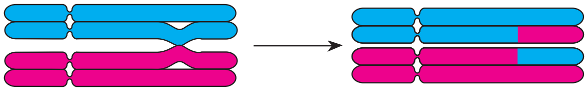


**Figure 3.21** Crossing over in a tetrad during prophase of meiosis I. Maternal chromosomes are *red;* paternal chromosomes are *blue.* Note the exchange of chromosome pieces after the process is completed.

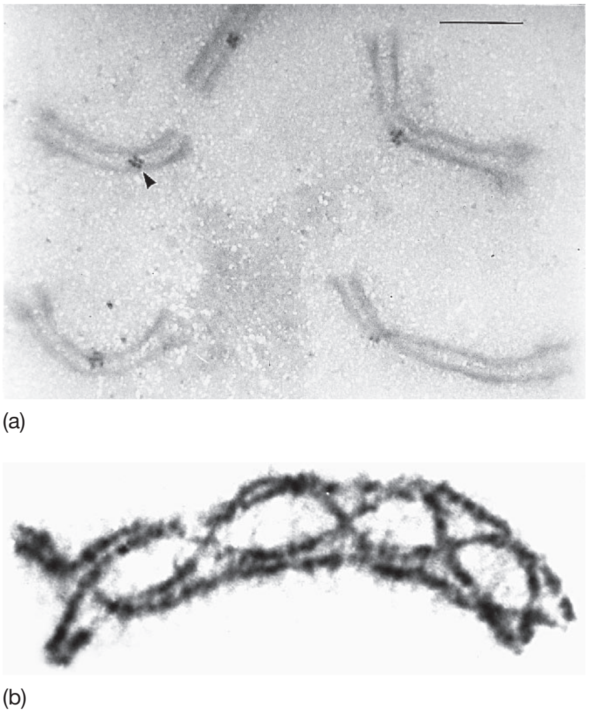
tromere. Crossing over can greatly increase the genetic variability in gametes by associating alleles that were not previously joined. (We examine the molecular mechanism of this process in chapter 12.) Before crossing over takes place, densely staining nodules are visible, first in zygonema and lasting through pachynema. These are called **recombination nodules** (fig. 3.22*a*); they are correlated with crossing over and presumably represent the enzymatic machinery present on the chromosomes.

As the chromosomes shorten and thicken further in diplonema, each chromosome can be seen to be made of two sister chromatids. Now the chromosome figures are referred to as **tetrads** because each is made up of four chromatids (see fig. 3.19). At about this time, the synaptonemal complex disintegrates in all but the areas of the **chiasmata** (singular: chiasma), the X-shaped configurations marking the places of crossing over (fig. 3.22*b*). Virtually all tetrads exhibit chiasmata; in cases in which no crossing over occurs, the tetrads tend to fall apart and segregate randomly. Thus, crossing over not only increases genetic diversity but also ensures the proper separation of homologous chromosomes. A meiosis-specific form of cohesin keeps sister chromatids together.

During the diplotene stage, chromosomes can again uncondense and become active. This is especially obvious in amphibians and birds, which produce a great amount of cytoplasmic nutrient for the future zygote. Recondensation of the chromosomes takes place at the end of diplonema. This stage can be very long; in human females, it begins in the fetus and does not complete until the egg is shed during ovulation, sometimes more than fifty years later. As prophase I moves into diakinesis, the chromosomes become very condensed (see fig. 3.19).

## Metaphase I and Anaphase I

Metaphase I is marked by the breakdown of the nuclear membrane and the attachment of kinetochore microtubules to the tetrads. Unlike in mitosis, in which sister chromatids are pulled apart because each sister kinetochore is attached to a different pole, both sister kinetochores become attached to spindle microtubules coming from the same pole in metaphase I (fig. 3.23). During anaphase I, cohesin breaks down every place but at the centromeres, allowing sister chromatids to be pulled to the same pole: homologous chromosomes are separated (fig. 3.24). This meiotic division is therefore called a **reductional division** because it reduces the number of chromosomes to half the diploid number in each daughter cell. For every tetrad there is now one chromosome in the form of a chromatid pair, known as a **dyad** or **monovalent,** at each pole of the cell. The initial objective of meiosis, separating homologues into different daughter cells, is accomplished. However, since each dyad consists of two sister chromatids, a second, mitosis-like division is required to reduce each chromosome to a single chromatid.



(a)

(b)

**Figure 3.22** (*a*) Recombination nodules (*arrowhead*) in spermatocytes of the pigeon, *Columba livia.* (Bar = 1 μm.) (*b*) A tetrad from the grasshopper, *Chorthippus parallelus,* at diplonema with five chiasmata. ([a] From M. I. Pigozzi and A. J. Solari, "Recombination Nodule Mapping and Chiasma Distribution in Spermatocytes of the Pigeon, *Columba livia,*" in *Genome,* 42: 308–314, 1999. Reprinted by permission. [b] Courtesy of Bernard John.)

## Telophase I and Prophase II

Depending on the organism, telophase I may or may not be greatly shortened in time. In some organisms, all the expected stages take place; chromosomes enter an interphase configuration as cytokinesis takes place. However, no chromosome duplication (DNA replication) occurs during this abbreviated interphase, termed **interkinesis.** Next, in these organisms, prophase II begins and

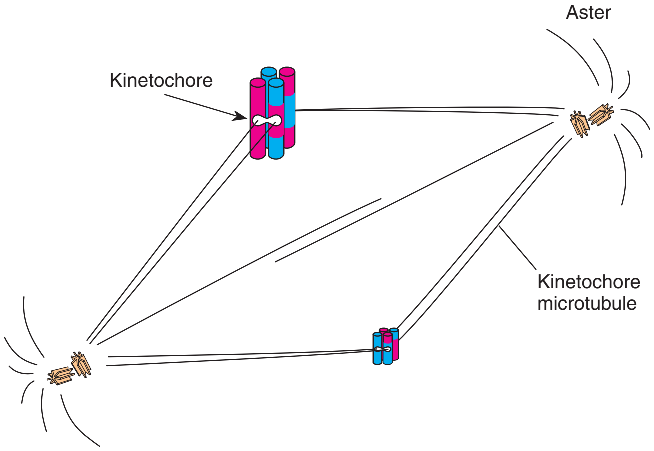


**Figure 3.23** Metaphase of meiosis I; $2n = 4$. Maternal chromosomes are *red;* paternal chromosomes are *blue.* Sister kinetochores (effectively single, merged kinetochores) are attached to microtubules from the same pole.

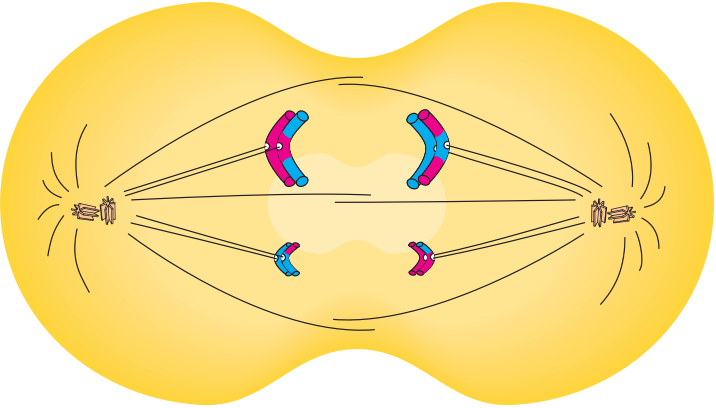


**Figure 3.24** Anaphase of meiosis I; $2n = 4$. Maternal chromosomes are *red;* paternal chromosomes are *blue.* Homologous chromosomes separate and move to opposite poles.

meiosis II proceeds. In still other organisms, the late anaphase I chromosomes go almost directly into metaphase II, virtually skipping telophase I, interphase, and prophase II.

## *Meiosis II*

Meiosis II is basically a mitotic division in which the chromatids of each chromosome are pulled to opposite poles. For each original cell entering meiosis I, four cells emerge at telophase II. Meiosis II is an **equational division;** although it reduces the amount of genetic material per cell by half, it does not further reduce the chromosome number per cell (fig. 3.25). (Sometimes it is simpler to concentrate on the behavior of centromeres during meiosis than on the chromosomes and chromatids. Meiosis I separates maternal from paternal centromeres, and meiosis II separates sister centromeres.) Figure 3.26 summarizes meiosis in corn (*Zea mays*).

In terms of chromosomes, meiosis begins with a diploid cell and produces four haploid cells. In terms of DNA, the process is a bit more complex but has the same

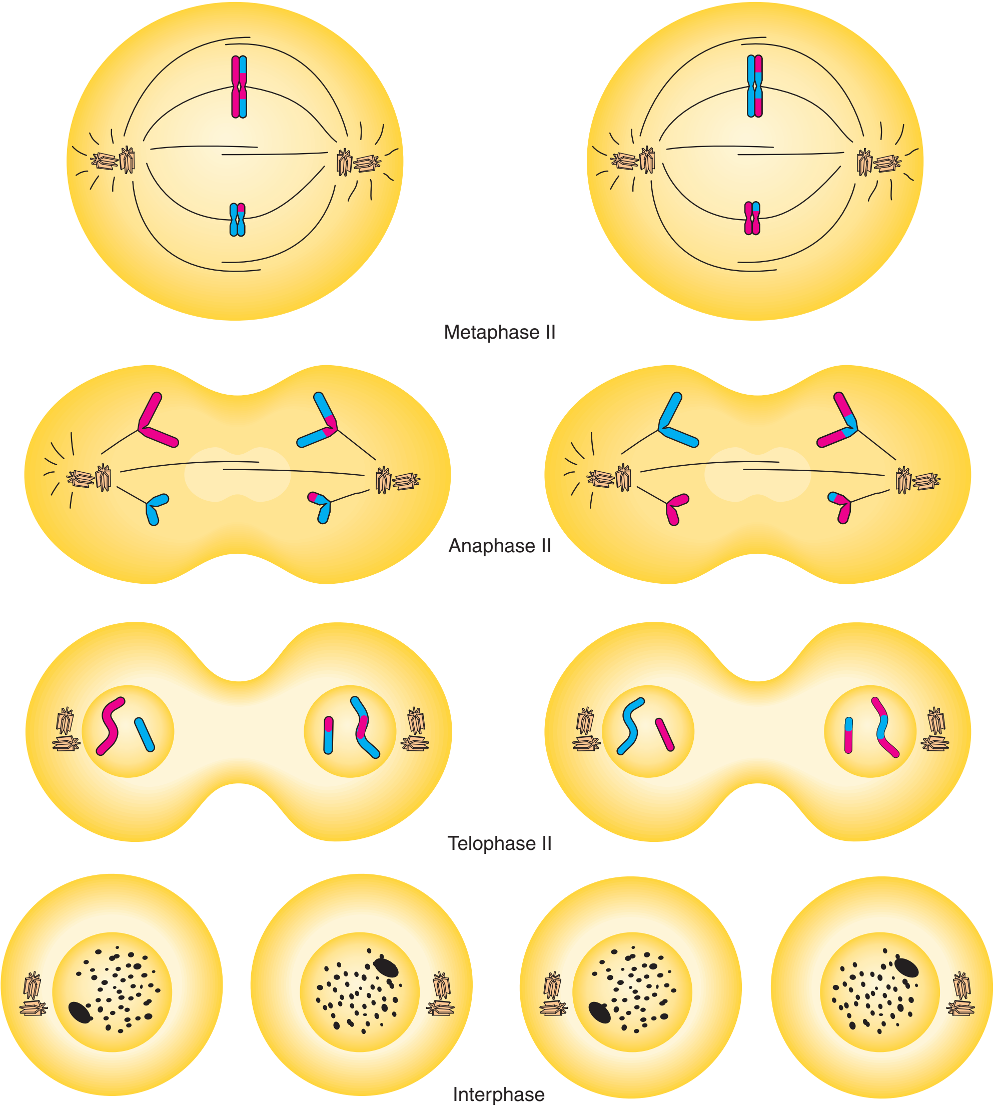


**Figure 3.25** Meiosis II; $2n = 4$. Maternal chromosomes are *red;* paternal chromosomes are *blue.*

result. Let us call the quantity of DNA in a gamete "C." A diploid cell before S phase has 2C DNA, and the same cell after S phase, but before mitosis, has 4C DNA. Mitosis reduces the quantity of DNA to 2C. A cell entering meiosis also has 4C DNA. After the first meiotic division, each daughter cell has 2C DNA, and after the second meiotic division, each daughter cell has C DNA, the quantity appropriate for a gamete.

## *The Significance of Meiosis*

Meiosis is significant for several reasons. First, it reduces the diploid number of chromosomes so that each of four daughter cells has one complete haploid chromosome set. Second, because of the randomness of the process of chromosomal separation, a very large number of different chromosomal combinations can form in the gametes. For example, in human beings, if each gamete could get either the maternal or paternal chromosome, and we have twenty-three chromosomal pairs, $2^{23}$ or 8,388,608 different combinations can occur. Third, because of crossing over, even more allelic combinations are possible. The process of creating new arrangements, either by crossing over or by independent segregation of homologous pairs of chromosomes, is called **recombination.** Assuming one hundred thousand genes in a human

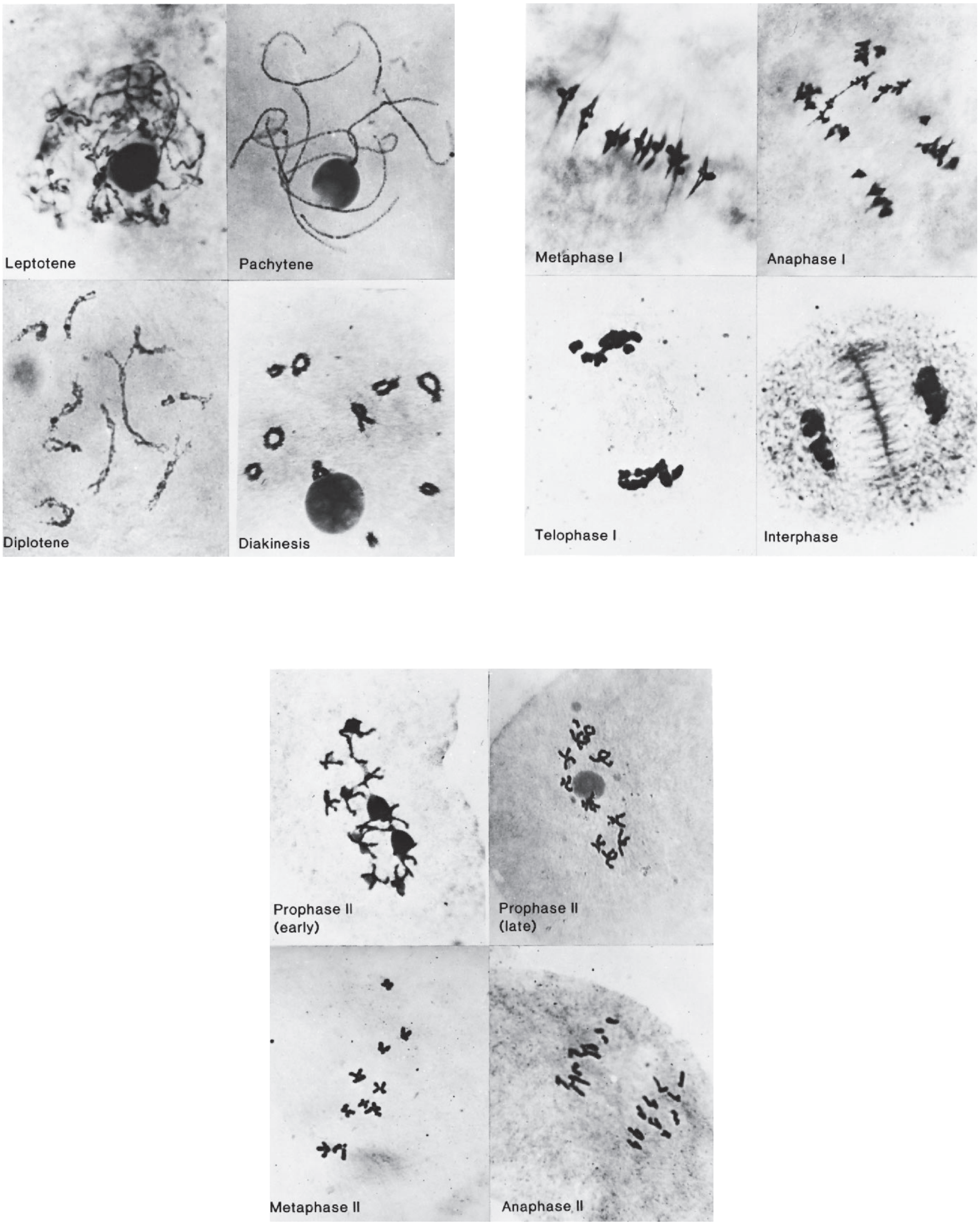

Leptotene
Pachytene
Diplotene
Diakinesis
Metaphase I
Anaphase I
Telophase I
Interphase
Prophase II
(early)
Prophase II
(late)
Metaphase II
Anaphase II

**Figure 3.26** Meiosis in corn (*Zea mays*). (Courtesy of Dr. M. M. Rhoades. "Meiosis in maize," *Journal of Heredity*, 41: 59–67, 1950. Reproduced by permission.)

being with two alleles each, $2^{100,000}$ different gametes could potentially arise by meiosis.

The behavior of any tetrad follows the pattern of Mendel's rule of segregation. At spore or gamete formation (meiosis), the diploid number of chromosomes is halved; each gamete receives only one chromosome from a homologous pair. This process, of course, explains Mendel's rule of segregation. Chromosomal behavior at meiosis also explains independent assortment (fig. 3.27). In anaphase I, the direction of separation is independent in different tetrads. Whereas one pole may get the maternal centromere from chromosomal pair number 1, it could get either the maternal or the paternal centromere from chromosomal pair number 2, and so on (see fig. 3.27). Alleles of one gene segregate independently of alleles of other genes. Very shortly after the rediscovery of Mendel's principles in 1900, geneticists were quick to realize this.

## MEIOSIS IN ANIMALS

In male animals, each meiosis produces four equal-sized **sperm cells** in a process called **spermatogenesis** (fig. 3.28). In vertebrates, a cell type in the testes known as a **spermatogonium** produces **primary spermatocytes,** as well as additional spermatogonia, by mitosis.

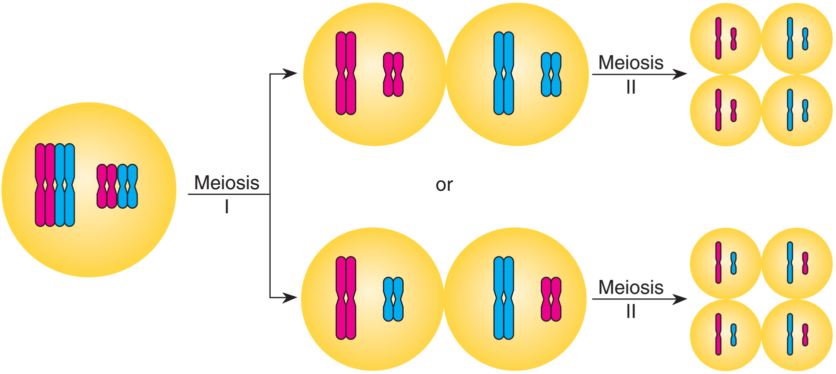

Meiosis
I
Meiosis
II
or
Meiosis
II

**Figure 3.27** Relationship of meiosis to the rule of independent assortment. Maternal (*red*) and paternal (*blue*) chromosomes separate independently in different tetrads.

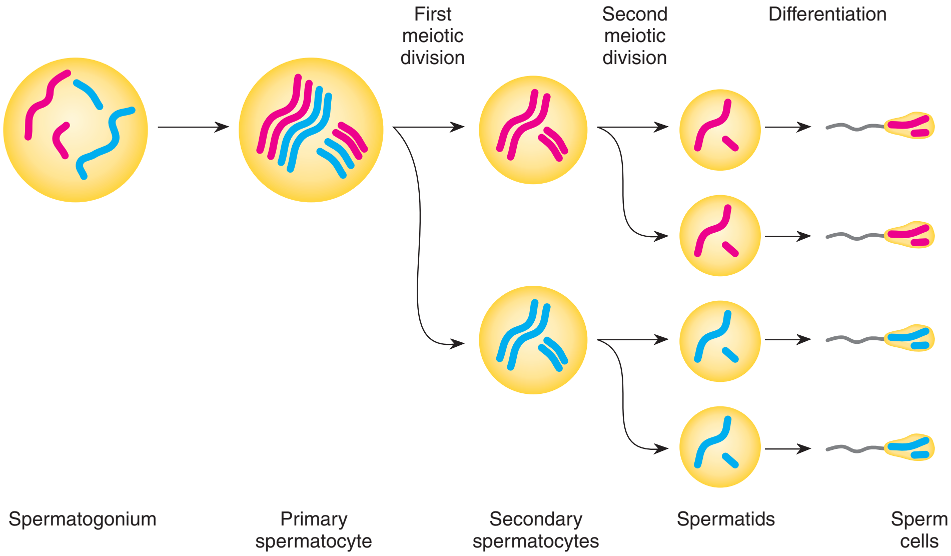

First
meiotic
division
Second
meiotic
division
Differentiation
Spermatogonium
Primary
spermatocyte
Secondary
spermatocytes
Spermatids
Sperm
cells

**Figure 3.28** Spermatogenesis; $2n = 4$. Maternal chromosomes are *red;* paternal chromosomes are *blue.*

The primary spermatocytes then undergo meiosis. After the first meiotic division, these cells are known as **secondary spermatocytes;** after the second meiotic division, they are known as **spermatids.** The spermatids mature into spermatozoa by a process called **spermiogenesis**—with four sperm cells resulting from each primary spermatocyte. In human beings and other vertebrates without a specific mating season, the process of spermatogenesis is continuous throughout adult life. A normal human male may produce several hundred million sperm cells per day.

During embryonic development in human females, cells in the ovary, known as **oogonia,** proliferate by numerous mitotic divisions to form **primary oocytes.** About one million form per ovary. These begin the first meiotic division and then stop before the birth of the female in a prolonged diplonema, called the **dictyotene** stage. A primary oocyte does not resume meiosis until the female is past puberty, when, under hormonal control, ovulation takes place. This process usually occurs for only one oocyte per month during the female's reproductive life span (from about twelve to fifty years of age). Meiosis only then proceeds in the ovulated oocyte. In the female, the two cells formed by meiosis I are of unequal size. One, termed the **secondary oocyte,** contains almost all the nutrient-rich cytoplasm; the other, a **polar body,** receives very little cytoplasm. The second meiotic division in the larger cell yields another polar body and an **ovum.** The first polar body may or may not divide to form two other polar bodies, which ultimately disintegrate. Thus, **oogenesis** produces cells of unequal size—an ovum and two or three polar bodies (fig. 3.29). Cells of unequal size are produced because the oocyte nucleus and meiotic spindle reside very close to the surface of this large cell.

## LIFE CYCLES

For eukaryotes, the basic pattern of the life cycle alternates between a diploid and a haploid state (see fig. 3.1). With the exception of the life cycles of bacteria and viruses, all life cycles are modifications of this general pattern. Bacteria, including blue-green algae, have a single circular chromosome; with exceptions described later, they are always in the haploid state. They divide by replicating their DNA and having the two copies separate into two daughter cells by simple cell division (see chapter 7). Viruses, on the border of being called alive, insert their genetic material into the cells of other organisms, and then manufacture new copies of themselves (see chapter 7).

Most animals are diploids that form gametes by meiosis, then restore the diploid number by fertilization. Exceptions, however, are numerous. For example, in the bees, wasps, and ants (hymenoptera), males are haploid and produce gametes by mitosis; females are diploid. Some fishes exist by **parthenogenesis,** in which the offspring come from unfertilized eggs that do not undergo meiosis. And, in some copepods, the sexual and parthenogenetic stages of their life cycles alternate.

The general pattern of the life cycle of plants alternates between two distinct generations, each of which, depending on the species, may exist independently. In lower plants, the haploid generation predominates, whereas in higher plants, the diploid generation is dominant. In flowering plants **(angiosperms),** the plant you see is the diploid **sporophyte** (see fig. 3.1). It is referred to as a sporophyte because, through meiosis, it will give rise to spores. The spores germinate into the alternate generation, the haploid **gametophyte,** which produces gametes by mitosis. Fertilization then produces the next generation

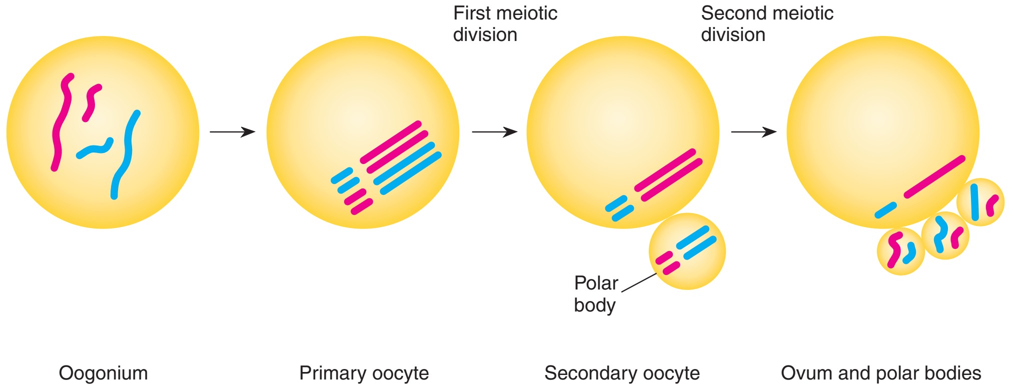


**Figure 3.29** Oogenesis; $2n = 4$. Maternal chromosomes are *red;* paternal chromosomes are *blue.*

of diploid sporophytes. In lower plants, the gametophyte has an independent existence; in angiosperms, this generation is radically reduced. For example, in corn (fig. 3.30), an angiosperm, the mature corn plant is the sporophyte. The male flowers produce microspores by meiosis. After mitosis, three cells exist in each spore, a structure that we call a **pollen grain,** the male gametophyte. In female flowers, meiosis produces megaspores. Mitosis within a megaspore produces an embryo sac of seven cells with eight nuclei. This is the female gametophyte. A sperm cell fertilizes the egg cell. The two polar nuclei of the embryo sac are fertilized by a second sperm cell, producing triploid ($3n$) nutritive endosperm tissue. The sporophyte grows from the diploid fertilized egg.

Many fungi and protista are haploid. Fertilization produces a diploid stage, which almost immediately undergoes meiosis to form haploid cells. These cells, in turn, increase in number by mitosis. We will analyze organisms such as *Neurospora,* the pink bread mold, in more detail later (see chapter 6).

Much of our knowledge of genetics derives from the study of specific organisms with unique properties. Mendel found pea plants useful because he could control matings carefully, their generation time was only a year, he could easily grow them in his garden, and they had the discrete traits that he was seeking. Our interest in human beings is obvious. However, we are members of a very difficult species to study experimentally. We have a long generation time and a small number of offspring from matings that we cannot tailor for research purposes. The fruit fly, ***Drosophila melanogaster,*** is one of the organisms geneticists have studied most extensively. Fruit flies have a short generation time (twelve to fourteen days), which means that many matings can be carried out in a reasonable amount of time. In addition, they do exceptionally well in the laboratory, they have many easily observable mutants, and in several organs they have giant banded chromosomes of great interest to cytogeneticists.

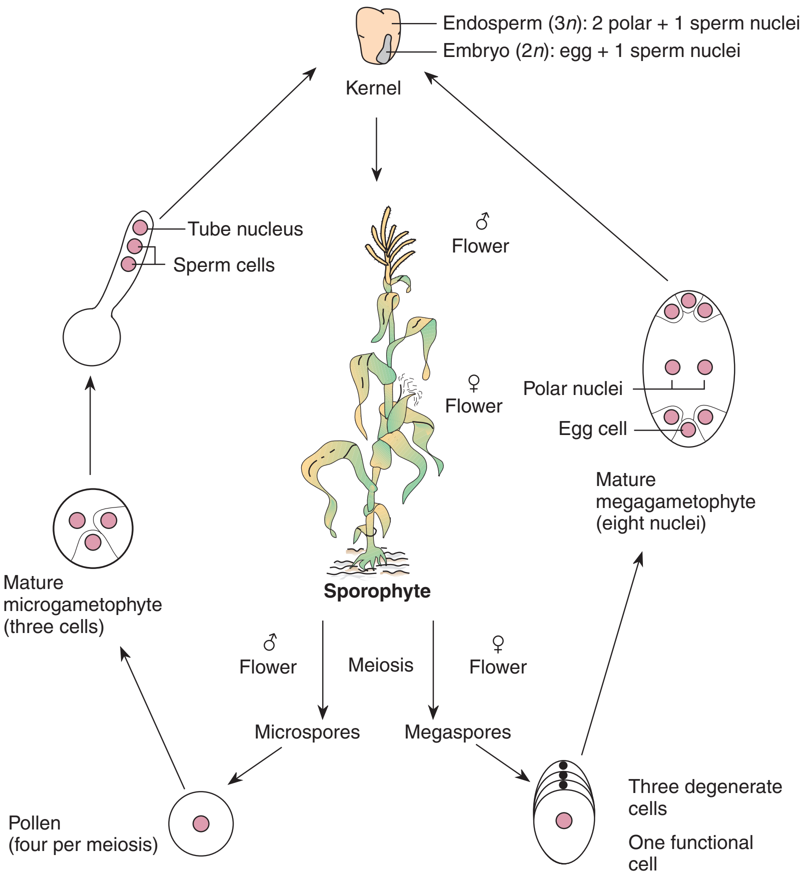

Endosperm (3n): 2 polar + 1 sperm nuclei
Embryo (2n): egg + 1 sperm nuclei
Kernel
Tube nucleus
Sperm cells
♂ Flower
♀ Flower
Polar nuclei
Egg cell
Mature megagametophyte (eight nuclei)
Mature microgametophyte (three cells)
Sporophyte
♂ Flower
Meiosis
♀ Flower
Microspores
Megaspores
Three degenerate cells
One functional cell
Pollen (four per meiosis)

**Figure 3.30** Life cycle of the corn plant.

Note that species used in food production tend to be intermediate in their life cycles. That is, many crop plants, such as peas and corn, have only one generation interval per year under normal circumstances. (We use the term *generation interval* here in the broadest sense, as the time it takes to complete an entire life cycle; see also chapter 19.) Crop plants are easier to work with from a genetic standpoint than people, but much more difficult than, say, *Drosophila* or bacteria (table 3.4). Because of their relatively long generation interval, crop plants are limited in their utility for studying basic genetic concepts or applying genetic technology to agriculture.

As you make your way through this book and through other readings on genetics, and as you come across studies involving new organisms, ask yourself the question, What are the properties of this organism that make it ideal for this type of research?

## CHROMOSOMAL THEORY OF HEREDITY

In a paper in 1903, cytologist Walter Sutton firmly stated the concepts we have developed here: The behavior of chromosomes during meiosis explains Mendel's principles. Genes, then, must be located on chromosomes. This idea, which several other biologists were also developing at the time, was immediately accepted, ushering in the era of the **chromosomal theory of inheritance.** During this era, intensive effort was devoted to studying the relationships between genes and chromosomes. The major portion of the first section of this book is devoted to classical studies of **linkage** and **mapping.** Linkage deals with the association of genes to each other and to specific chromosomes. Mapping deals with the sequence of genes on a chromosome and the distances between genes on the same chromosome. This is basic information for a study of the structure and function of genes. Here we introduce a new term for the gene. The term **locus** (plural: *loci*), meaning "place" in Latin, refers to the location of a gene on the chromosome.

**Table 3.4** Approximate Generation Intervals of Some Organisms of Genetic Interest

| Organism | Approximate Generation Interval |
|---|---|
| Intestinal bacterium (*Escherichia coli*) | 20 minutes |
| Bacterial virus (*lambda*) | 1 hour |
| Pink bread mold (*Neurospora crassa*) | 2 weeks |
| Fruit fly (*Drosophila melanogaster*) | 2 weeks |
| House mouse (*Mus musculus*) | 2 months |
| Corn (*Zea mays*) | 6 months |
| Sheep (*Ovus aries*) | 1 year |
| Cattle (*Bos taurus*) | 2 years |
| Human being (*Homo sapiens*) | 14 years |

## SUMMARY

**STUDY OBJECTIVE 1:** To observe the morphology of chromosomes 48–50

Chromosomes are made of chromatin and divided by centromeres. Within centromeres are kinetochores, attachment points for spindle fibers. Structure within the chromosomes is evident from bands on chromosomes called chromomeres.

**STUDY OBJECTIVE 2:** To understand the processes of mitosis and meiosis 50–61

During eukaryotic cell division, the processes of mitosis and meiosis apportion the chromosomes to daughter cells. Both processes are preceded by chromosome replication during the S phase of the cell cycle, which is under genetic control. In mitosis, the two sister chromatids making up each replicated chromosome separate into two daughter cells. Sex cells—gametes in animals and spores in plants—are produced by the two-stage process of meiosis. In meiosis, homologous chromosomes are first separated into two daughter cells, and then the sister chromatids making up each chromosome are distributed to two new daughter cells. We end up with four cells, each with the haploid chromosomal complement. The spindle is the apparatus that separates chromosomes in both mitosis and meiosis.

**STUDY OBJECTIVE 3:** To analyze the relationships between meiosis and Mendel's rules 61–65

The behavior of chromosomes during meiosis explains Mendel's two principles, segregation and independent assortment.

At the end of this chapter, we define the chromosomal theory of inheritance, the concept that shapes the first section of this book. This theory states that genes are located on chromosomes; their positions and order on the chromosomes can be discovered by mapping techniques described in later chapters.

## SOLVED PROBLEMS

**PROBLEM 1:** What are the differences between chromosomes and chromatids?

***Answer:*** In higher organisms, a chromosome is a linear DNA molecule complexed with protein and, generally, with a centromere somewhere along its length. During the cell cycle, in the S phase, the DNA replicates and each chromosome is duplicated. The duplication is visible in the early stages of mitosis and meiosis when chromosomes shorten. At this point, each duplicated chromosome is made up of two chromatids. The chromatids are called chromosomes when their centromeres are pulled to opposite poles of the spindle and each chromatid becomes independent.

**PROBLEM 2:** What are the relationships between mitosis and meiosis and Mendel's rules of segregation and independent assortment?

***Answer:*** The process of mitosis does not relate directly to Mendel's rules. The behavior of chromosomes during meiosis, however, explains both segregation and independent assortment. Segregation is explained by the fact that only one chromosome from each homologous pair goes into a gamete; this is also true for the maternal and paternal alleles of a given gene. Independent assortment is explained by the independent behavior of each tetrad at meiosis. That is, the separation of maternal and paternal alleles in one tetrad is independent of the separation of maternal and paternal alleles in any other tetrad.

**PROBLEM 3:** A hypothetical organism has six chromosomes ($2n = 6$). How many different combinations of maternal and paternal chromosomes can appear in the gametes?

***Answer:*** You could figure this empirically by listing all combinations. For example, let A, B, and C = maternal chromosomes and A′, B′, and C′ = paternal chromosomes. Two combinations in the gametes could be A B C′ and A′ B′ C; obviously, several other combinations exist. It is easier to recall that $2^n$ = number of combinations, where $n$ = the number of chromosome pairs. In this case, $n = 3$, so we expect $2^3 = 8$ different combinations.

## EXERCISES AND PROBLEMS*

### CHROMOSOMES

**1.** What are the major differences between prokaryotes and eukaryotes?

**2.** What is the difference between a centromere and a kinetochore?

**3.** What is the difference between sister and nonsister chromatids? Between homologous and nonhomologous chromosomes?

**4.** In human beings, $2n = 46$. How many chromosomes would you find in a
- **a.** brain cell?
- **b.** red blood cell?
- **c.** polar body?
- **d.** sperm cell?
- **e.** secondary oocyte?

(See also MEIOSIS IN ANIMALS)

### MITOSIS

**5.** You are working with a species with $2n = 6$, in which one pair of chromosomes is telocentric, one pair subtelocentric, and one pair metacentric. The *A, B,* and *C* loci, each segregating a dominant and recessive allele (*A* and *a, B* and *b, C* and *c*), are each located on different chromosome pairs. Draw the stages of mitosis.

**6.** Identify stages *a–f* in the nuclear division shown in figure 1 (on the next page). Include the process, stage, and diploid number (e.g., meiosis I, prophase, $2n = 10$). Keep in mind that one picture could represent more than one process and stage. Chromosomes are drawn as threads, with circles representing kinetochores. (See also MEIOSIS)

**7.** When during the cell cycle does chromosome replication take place?

**8.** A mature human sperm cell has $c$ amount of DNA. How much DNA ($c$, $2c$, $4c$, etc.) will a somatic cell have if it is in
- **a.** $G_1$?
- **b.** $G_2$?

How much DNA will be in a cell at the end of meiosis I?

---

*Answers to selected exercises and problems are on page A-3.

### MEIOSIS

**9.** Given the same information as in problem 5, diagram one of the possible meioses. How many different gametes can arise, absent crossing over? What variation in gamete genotype is introduced by a crossover between the *A* locus and its centromere?

**10.** How many bivalents, tetrads, and dyads would you find during meiosis in human beings? in fruit flies? in the other species of table 3.3?

**11.** Can you devise a method of chromosome partitioning during gamete formation that would not involve synapsis—that is, can you reengineer meiosis without passing through a synapsis stage?

**12.** What are the differences between a reductional and an equational division? What do these terms refer to?

**13.** How does the process of meiosis explain Mendel's two rules of inheritance?

**14.** *Drosophila* has four pairs of chromosomes. Let chromosomes from the male parent be A, B, C, and D, and those from the female parent be A′, B′, C′, and D′. What fraction of the gametes from an AA′ BB′ CC′ DD′ individual will be

- **a.** all of paternal origin?
- **b.** all of maternal origin?
- **c.** half of maternal origin and half of paternal origin?

**15.** Wheat has $2n = 42$ and rye has $2n = 14$ chromosomes. Explain why a wheat-rye hybrid is usually sterile.

**16.** The arctic fox has fifty small chromosomes, and the red fox has thirty-eight larger chromosomes. Hybrids of these two species are sterile, but cytological studies during meiosis in these hybrids reveal both paired and unpaired chromosomes.

- **a.** Account for the sterility of the hybrids.
- **b.** How can you explain the paired chromosomes?

**17.** An organism has six pairs of chromosomes. In the absence of crossing over, how many different chromosomal combinations are possible in the gametes?

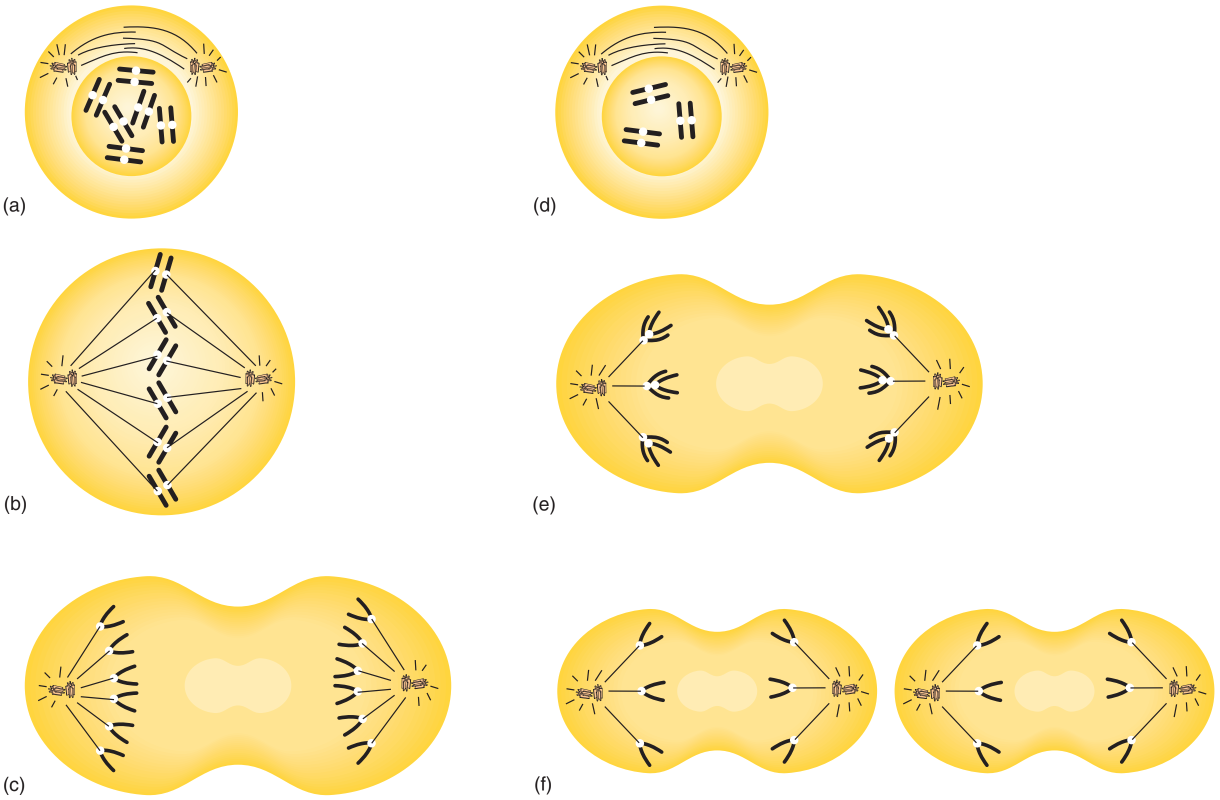


**Figure 1** Stages in nuclear division.

### MEIOSIS IN ANIMALS

**18.** How many sperm come from ten primary spermatocytes? How many ova from ten primary oocytes?

**19.** How do the quantity of genetic material and the ploidy change from stage to stage of spermatogenesis and oogenesis (see figs. 3.28 and 3.29)? (Consider the spermatogonium and the oogonium to be diploid, with the chromosome number arbitrarily set at two.)

**20.** How many sperm cells will form from
- **a.** fifty primary spermatocytes?
- **b.** fifty secondary spermatocytes?
- **c.** fifty spermatids?

**21.** In human beings, how many eggs will form from
- **a.** fifty primary oocytes?
- **b.** fifty secondary oocytes?

### LIFE CYCLES

**22.** In corn (see fig. 3.30), the diploid number is twenty. How many chromosomes would you find in a(n)
- **a.** sporophyte leaf cell?
- **b.** embryo cell?
- **c.** endosperm cell?
- **d.** pollen grain?
- **e.** polar nucleus?

**23.** If a dihybrid corn plant is self-fertilized, what genotypes of the triploid endosperm can result? If you know the endosperm genotype, can you determine the genotype of the embryo?

**24.** Change the generalized life cycles of figure 3.1 so they describe the life cycles of human beings, peas, and *Neurospora.*

**25.** If the cytoplasm rather than nuclear genes controlled inheritance, what might be the relationship in phenotype and genotype between an organism and its parents in
- **a.** *Drosophila*?
- **b.** corn?
- **c.** *Neurospora*?

**26.** A drone (male) honeybee is haploid (arising from unfertilized eggs), and a queen (female) is diploid. Draw a testcross between a dihybrid queen and a drone. How many different kinds of sons and daughters might result from this cross?

**27.** The plant *Arabidopsis thaliana* has five pairs of chromosomes: AA, BB, CC, DD, and EE. If this plant is self-fertilized, what chromosome complement would be found in a root cell of the offspring?
- **a.** A B C D E
- **b.** AA BB CC DD EE
- **c.** AAA BBB CCC DDD EEE
- **d.** AAAA BBBB CCCC DDDD EEEE

**28.** In wheat, the haploid number is twenty-one. How many chromosomes would you expect to find in
- **a.** the tube nucleus?
- **b.** a leaf cell?
- **c.** the endosperm?

### CHROMOSOMAL THEORY OF HEREDITY

**29.** A hypothetical organism has two distinct chromosomes ($2n = 4$) and fifty known genes, each with two alleles. If an individual is heterozygous at all known loci, how many gametes can be produced if
- **a.** all genes behave independently?
- **b.** all genes are completely linked?

## CRITICAL THINKING QUESTIONS

**1.** Can meiosis occur in a haploid cell? can mitosis?

**2.** What is the minimum number of chromosomes that an organism can have? the maximum number?

*Suggested Readings for chapter 3 are on page B-1.*

# 4

# PROBABILITY AND STATISTICS

## STUDY OBJECTIVES

1. To understand the rules of probability and how they apply to genetics 71
2. To understand the use of the chi-square statistical test in genetics 74

## STUDY OUTLINE

**Probability** 71
  Types of Probabilities 71
  Combining Probabilities 71
**Use of Rules** 72
**Statistics** 74
  Hypothesis Testing 74
  Chi-Square 76
  Failing to Reject Hypotheses 77
*Summary* 78
*Solved Problems* 78
*Exercises and Problems* 79
*Critical Thinking Questions* 81

An agricultural worker studies variability in plants in a greenhouse. Probability influences the differences among organisms. (© David Joel/Tony Stone Images.)

In an experimental science, such as genetics, scientists make decisions about hypotheses on the basis of data gathered during experiments. Geneticists must therefore have an understanding of probability theory and statistical tests of hypotheses. Probability theory allows geneticists to construct accurate predictions of what to expect from an experiment. Statistical testing of hypotheses, particularly with the chi-square test, allows geneticists to have confidence in their interpretations of experimental data.

## PROBABILITY

Part of Gregor Mendel's success was due to his ability to work with simple mathematics. He was capable of turning numbers into ratios and deducing the mechanisms of inheritance from them. Taking numbers that did not exactly fit a ratio and rounding them off to fit lay at the heart of Mendel's deductive powers. The underlying rules that make the act of "rounding to a ratio" reasonable are the rules of probability.

In the **scientific method,** scientists make predictions, perform experiments, and gather data that they then compare with their original predictions (see chapter 1). However, even if the bases for the predictions are correct, the data almost never exactly fit the predicted outcome. The problem is that we live in a world permeated by random, or **stochastic,** events. A bright new penny when flipped in the air twice in a row will not always give one head and one tail. In fact, that penny, if flipped one hundred times, could conceivably give one hundred heads. In a stochastic world, we can guess how often a coin should land heads up, but we cannot know for certain what the next toss will bring. We can guess how often a pea should be yellow from a given cross, but we cannot know with certainty what the next pod will contain. Thus, we need **probability theory** to tell us what to expect from data. This chapter closes with some thoughts on statistics, a branch of mathematics that helps us with criteria for supporting or rejecting our hypotheses.

### *Types of Probabilities*

The **probability** ($P$) that an event will occur is the number of favorable cases ($a$) divided by the total number of possible cases ($n$):

$$P = a/n$$

The probability can be determined either by observation (empirical) or by the nature of the event (theoretical). For example, we observe that about one child in ten thousand is born with phenylketonuria. Therefore, the probability that the next child born will have phenylketonuria is 1/10,000. The odds based on the geometry of an event are, for example, like the familiar toss of dice. A die (singular of dice) has six faces. When that die is tossed, there is no reason one face should land up more often than any other. Thus, the probability of any one of the faces being up (e.g., a four) is one-sixth:

$$P = a/n = 1/6$$

Similarly, the probability of drawing the seven of clubs from a deck of cards is

$$P = 1/52$$

The probability of drawing a spade from a deck of cards is

$$P = 13/52 = 1/4$$

The probability (assuming a 1:1 sex ratio, though the actual ratio is about 1.06 males per female born in the United States) of having a daughter in any given pregnancy is

$$P = 1/2$$

And the probability that an offspring from a self-fertilized dihybrid will show the dominant phenotype is

$$P = 9/16$$

From the probability formula, we can say that an event with certainty has a probability of one, and an event that is an impossibility has a probability of zero. If an event has the probability of $P$, all the other alternatives combined will have a probability of $Q = 1 - P$; thus $P + Q = 1$. That is, the probability of the completely dominant phenotype in the $F_2$ of a selfed dihybrid is 9/16. The probability of any other phenotype is 7/16, and when the two are added together, they equal 16/16, or 1.

### *Combining Probabilities*

The basic principle of probability can be stated as follows: If one event has $c$ possible outcomes and a second event has $d$ possible outcomes, then there are $cd$ possible outcomes of the two events. From this principle, we obtain three rules that concern us as geneticists.

To understand these rules of probability requires a few definitions. *Mutually exclusive events* are events in which the occurrence of one possibility excludes the occurrence of the other possibilities. In the throwing of a die, for example, only one face can land up. Thus, if it comes up a four, it precludes the possibility of any of the other faces. Similarly, a blue-eyed daughter is mutually exclusive of a brown-eyed son or any other combination of gender and eye color. *Independent events,* however, are events whose outcomes do not influence one another. For example, if two dice are thrown, the face value of one die is not able to affect the face value of the other; they are thus independent of each other. Similarly, the gender of one child in a family is generally independent of the gender of the children who have come before or might come after. Finally, *unordered events* are events whose probability of outcome does not depend on the order in which the events occur; the probabilities combine both mutual exclusivity and independence. For example, when two dice (one red, one green) are thrown at the same time, we generally do not specify which die has which value; a seven can occur whether the green die is the four or the red die is the four. Similarly, the probability that a family of several children will have two boys and one girl is the same irrespective of their birth order. If the family has two boys and one girl, it does not matter whether the daughter is born first, second, or third. In general, probabilities differ depending on whether order is specified. With these definitions in mind, let us look at three rules of probability that affect genetics.

### *1. Sum Rule*

When events are mutually exclusive, the **sum rule** is used: The probability that one of several mutually exclusive events will occur is the sum of the probabilities of the individual events. This is also known as the *either-or rule.* For example, what is the probability, when we throw a die, of its showing *either* a four *or* a six? According to the sum rule,

$$P = 1/6 + 1/6 = 2/6 = 1/3$$

### *2. Product Rule*

When the occurrence of one event is independent of the occurrence of other events, the **product rule** is used: The probability that two independent events will both occur is the product of their separate probabilities. This is known as the *and rule.* For example, the probability of throwing a die two times and getting a four *and* then a six, in that order, is

$$P = 1/6 \times 1/6 = 1/36$$

### *3. Binomial Theorem*

The **binomial theorem** is used for unordered events: The probability that some arrangement will occur in which the final order is not specified is defined by the binomial theorem. For example, what is the probability when tossing two pennies simultaneously of getting a head and a tail? We will look more closely at how to use the rules of probability to answer this question.

## USE OF RULES

There are several ways to calculate the probability just asked for. To put the problem in the form for rule 3 is the quickest method, but this problem can also be solved by using a combination of rules 1 and 2. For each penny, the probability of getting a head (H) *or* a tail (T) is

$$\text{for H: } P = 1/2$$
$$\text{for T: } Q = 1/2$$

Tossing the pennies one at a time, it is possible to get a head *and* a tail in two ways:

first head, then tail (HT)
or
first tail, then head (TH)

Within a sequence (HT or TH), the probabilities apply to independent events. Thus, the probability for any one of the two sequences involves the product rule (rule 2):

$$1/2 \times 1/2 = 1/4 \text{ for HT or TH}$$

The two sequences (HT or TH) are mutually exclusive. Thus, the probability of getting either of the two sequences involves the sum rule (rule 1):

$$1/4 + 1/4 = 1/2$$

Thus, for unordered events, we can obtain the probability by combining rules 1 and 2. The binomial theorem (rule 3) provides the shorthand method.

To use rule 3, we must state the theorem as follows: If the probability of an event ($X$) is $p$ and an alternative ($Y$) is $q$, then the probability in $n$ trials that event $X$ will occur $s$ times and $Y$ will occur $t$ times is

$$P = \frac{n!}{s!t!} p^s q^t$$

In this equation, $s + t = n$, and $p + q = 1$. The symbol !, as in $n$!, is called **factorial,** as in "$n$ factorial," and is the product of all integers from $n$ down to one. For example, $7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$. Zero factorial equals one, as does anything to the power of zero ($0! = n^0 = 1$).

Now, what is the probability of tossing two pennies at once and getting one head and one tail? In this case, $n = 2$, $s$ and $t = 1$, and $p$ and $q = 1/2$. Thus,

$$P = \frac{2!}{1!1!}(1/2)^1(1/2)^1 = 2(1/2)^2 = 1/2$$

This is, of course, our original answer. Now on to a few more genetically relevant problems. What is the probability that a family with six children will have five girls and one boy? (We assume that the probability of either a son or a daughter equals 1/2.) Since the order is not specified, we use rule 3:

$$P = \frac{6!}{5!1!}(1/2)^5(1/2)^1 = 6(1/2)^6 = 6/64 = 3/32$$

What would happen if we asked for a specific family order, in which four girls were born, then one boy, and then one girl? This would entail rule 2; for a sequence of six independent events:

$$P = 1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2 = 1/64$$

When no order is specified, the probability is six times larger than when the order is specified; the reason is that there are six ways of getting five girls and one boy, and the sequence 4-1-1 is only one of them. Rule 3 tells us that there are six ways. These are (letting B stand for boy and G for girl) as follows:

**Birth Order**

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| B | G | G | G | G | G |
| G | B | G | G | G | G |
| G | G | B | G | G | G |
| G | G | G | B | G | G |
| G | G | G | G | B | G |
| G | G | G | G | G | B |

Let us look at yet another problem. If two persons, heterozygous for albinism (a recessive condition), have four children, what is the probability that all four children will be normal? The answer is simply $(3/4)^4$ by rule 2. What is the probability that three will be normal and one albino? If we specify which of the four children will be albino (e.g., the fourth), then the probability is $(3/4)^3(1/4)^1 = 27/256$. If, however, we do not specify order,

$$P = \frac{4!}{3!1!}(3/4)^3(1/4)^1 = 4(3/4)^3(1/4)^1$$
$$= 4(27/256) = 108/256$$

This is precisely four times the ordered probability because the albino child could have been born first, second, third, or last.

The formula for rule 3 is the formula for the terms of the **binomial expansion.** That is, if $(p + q)^n$ is expanded (multiplied out), the formula $(n!/s!t!)p^sq^t$ gives the probability for any one of these terms, given that $p + q = 1$ and that $s + t = n$. Since there are $(n + 1)$ terms in the binomial, the formula gives the probability for the term numbered $(t + 1)$. Two bits of useful information come from recalling that rule 3 is in reality the binomial expansion formula. First, if you have difficulty calculating the term, you can use **Pascal's triangle** to get the coefficients:

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1

Pascal's triangle is a triangular array made up of coefficients in the binomial expansion. It is calculated by starting any row with a 1, proceeding by adding two adjacent terms from the row above, and then ending with a 1. For example, the next row would be

1, (1 + 5), (5 + 10), (10 + 10), (10 + 5), (5 + 1), 1
or 1, 6, 15, 20, 15, 6, 1

These numbers give us the combinations for any $p^sq^t$ term. That is, in our previous example, $n = 4$; so we use the $(n + 1)$, or fifth, row of Pascal's triangle. (The second number in any row of the triangle gives the power of the expansion, or $n$. Here, 4 is the second number in the row.) We were interested in the case of one albino child in a family of four children, or $p^3q^1$, where $p$ is the probability of the normal child (3/4) and $q$ is the probability of an albino child (1/4). Hence, we are interested in the $(t + 1)$—that is, the $(1 + 1)$—or the second term of the fifth row of Pascal's triangle, which will tell us the number of ways of getting a four-child family with one albino child. That number is 4. Thus, using Pascal's triangle, we see that the solution to the problem is

$$4(3/4)^3(1/4)^1 = 108/256$$

This is the same as the answer we obtained the conventional way.

The second advantage from knowing that rule 3 is the binomial expansion formula is that we can now generalize to more than two outcomes. The general form for the **multinomial expansion** is $(p + q + r + \ldots)^n$ and the general formula for the probability is

$$P = \frac{n!}{s!t!u! \ldots} p^sq^tr^u \ldots$$

where $s + t + u + \ldots = n$ and $p + q + r + \ldots = 1$ For example, our albino-carrying heterozygous parents may want an answer to the following question: If we have five children, what is the probability that we will have two normal sons, two normal daughters, and one albino son? (This family will have no albino daughters.) By rule 2, the probability of

$$\begin{aligned}\text{a normal son} &= (3/4)(1/2) = 3/8\\ \text{a normal daughter} &= (3/4)(1/2) = 3/8\\ \text{an albino son} &= (1/4)(1/2) = 1/8\\ \text{an albino daughter} &= (1/4)(1/2) = 1/8\end{aligned}$$

Thus:

$$\begin{aligned}P &= \frac{5!}{2!2!1!0!}(3/8)^2(3/8)^2(1/8)^1(1/8)^0\\ &= 30(3/8)^4(1/8)^1 = 30(3)^4/(8)^5 = 2{,}430/32{,}768\\ &= 0.074\end{aligned}$$

# STATISTICS

In one of Mendel's experiments, $F_1$ heterozygous pea plants, all tall, were self-fertilized. In the next generation ($F_2$), he recorded 787 tall offspring and 277 dwarf offspring for a ratio of 2.84:1. Mendel saw this as a 3:1 ratio, which supported his proposed rule of inheritance. In fact, is 787:277 "roundable" to a 3:1 ratio? From a brief discussion of probability, we expect some deviation from an exact 3:1 ratio (798:266), but how much of a deviation is acceptable? Would 786:278 still support Mendel's rule? Would 785:279 support it? Would 709:355 (a 2:1 ratio) or 532:532 (a 1:1 ratio)? Where do we draw the line? It is at this point that the discipline of statistics provides help.

We can never speak with certainty about stochastic events. For example, take Mendel's cross. Although a 3:1 ratio is expected on the basis of Mendel's hypothesis, chance could mean that the data yield a 1:1 ratio (532:532), yet the mechanism could be the one that Mendel suggested. In other words, we could flip an honest coin and get ten heads in a row. Conversely, Mendel could have gotten exactly a 3:1 ratio (798:266) in his $F_2$ generation, yet his hypothesis of segregation could have been wrong. The point is that any time we deal with probabilistic events there is some chance that the data will lead us to support a bad hypothesis or reject a good one. Statistics quantifies these chances. We cannot say with certainty that a 2.84:1 ratio represents a 3:1 ratio; we can say, however, that we have a certain degree of confidence in the ratio. Statistics helps us ascertain these **confidence limits.**

Statistics is a branch of probability theory that helps the experimental geneticist in three ways. First, part of statistics deals with **experimental design**. A bit of thought applied before an experiment may help the investigator design the experiment in the most efficient way. Although he did not know statistics, Mendel's experimental design was very good. The second way in which statistics is helpful is in summarizing data. Familiar terms such as *mean* and *standard deviation* are part of the body of descriptive statistics that takes large masses of data and reduces them to one or two meaningful values. We examine further some of these terms and concepts in the chapter on quantitative inheritance (chapter 18).

## Hypothesis Testing

The third way that statistics is valuable to geneticists is in the **testing of hypotheses:** determining whether to support or reject a hypothesis by comparing the data to the predictions of the hypothesis. This area is the most germane to our current discussion. For example, was the ratio of 787:277 really indicative of a 3:1 ratio? Since we know now that we cannot answer with an absolute yes, how can we decide to what level the data support the predicted 3:1 ratio?

Statisticians would have us proceed as follows. To begin, we need to establish how much variation to expect. We can determine this by calculating a **sampling distribution:** the frequencies with which various possible events could occur in a particular experiment. For example, if we self-fertilized a heterozygous tall plant, we would expect a 3:1 ratio of tall to dwarf plants among the progeny. (The 3:1 ratio is our hypothesis based on the assumption that height is genetically controlled by one locus with two alleles.) If we looked at the first four offspring, what is the probability we would see three tall and one dwarf plant? We can calculate the answer using the formula for the terms of the binomial expansion:

$$P = \frac{4!}{3!1!}(3/4)^3(1/4)^1 = 108/256 = 0.42$$

Similarly, we can calculate the probability of getting all tall (81/256 = 0.32), two tall and two dwarf (54/256 = 0.21), one tall and three dwarf (12/256 = 0.05), and all dwarf (1/256 = 0.004) in this first set of four. Table 4.1 shows this distribution, as well as the distributions for samples of eight and forty progeny. Figure 4.1 shows these distributions in graph form.

As sample sizes increase (from four to eight to forty in fig. 4.1), the sampling distribution takes on the shape of a smooth curve with a peak at the true ratio of 3:1 (75% tall progeny)—that is, there is a high probability of getting very close to the true ratio. However, there is some chance the ratio will be fairly far off, and a very small part of the time our ratio will be very far off. It is important to see that any ratio could arise in a given experiment even though the true ratio is 3:1. At what point do we decide that an experimental result is not indicative of a 3:1 ratio?

Statisticians have agreed on a convention. When all the frequencies are plotted, as in figure 4.1, we can treat the area under the curve as one unit, and we can draw lines to mark 95% of this area (fig. 4.2). Any ratios in-

**Table 4.1** Sampling Distribution for Sample Sizes of Four, Eight, and Forty, Given a 3:1 Ratio of Tall and Dwarf Plants

| $n = 4$ | | $n = 8$ | | $n = 40$ | |
|---|---|---|---|---|---|
| **No. Tall Plants** | **Probability*** | **No. Tall Plants** | **Probability*** | **No. Tall Plants** | **Probability*** |
| 4 | $\frac{81}{256} = 0.32$ | 8 | 0.10 | 40 | 0.00001 |
| | | 7 | 0.27 | 39 | 0.0001 |
| 3 | $\frac{108}{256} = 0.42$ | 6 | 0.31 | 38 | 0.0009 |
| | | 5 | 0.21 | | . . . |
| 2 | $\frac{54}{256} = 0.21$ | 4 | 0.09 | 30 | 0.14 |
| | | 3 | 0.02 | | . . . |
| 1 | $\frac{12}{256} = 0.05$ | 2 | 0.004 | 2 | $0.59 \times 10^{-20}$ |
| | | 1 | 0.0004 | 1 | $0.10 \times 10^{-21}$ |
| 0 | $\frac{1}{256} = 0.004$ | 0 | 0.00002 | 0 | $0.83 \times 10^{-24}$ |

* Probabilities are calculated from the binomial theorem.

probability = $(n!/s!t!)p^s q^t$

where $n$ = number of progeny observed
$s$ = number of progeny that are tall
$t$ = number of progeny that are dwarf
$p$ = probability of a progeny plant being tall (3/4)
$q$ = probability of a progeny plant being dwarf (1/4)

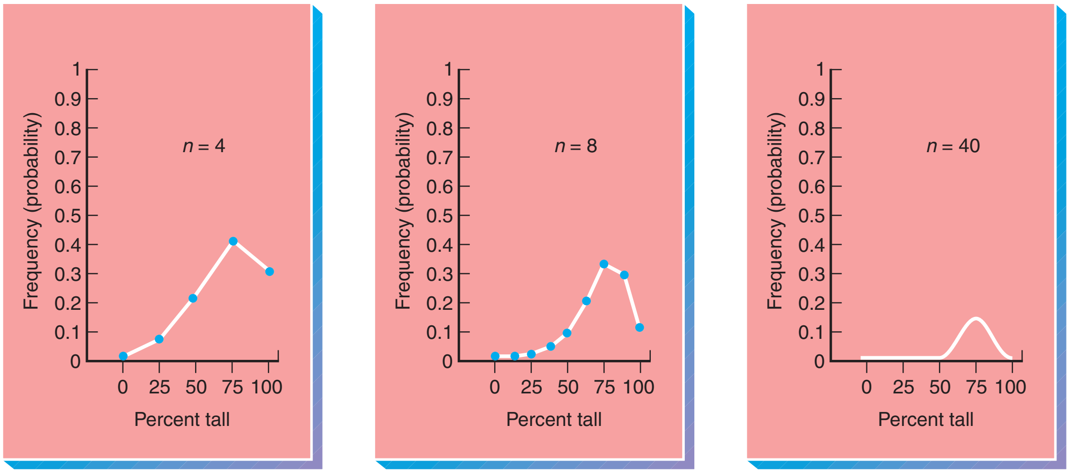


**Figure 4.1** Sampling distributions from an experiment with an expected ratio of three tall to one dwarf plant. As the sample size, *n,* gets larger, the distribution curve becomes smoother. These distributions are plotted terms of the binomial expansion (see table 4.1). Note also that as *n* gets larger, the peak of the curve gets lower because as more points (possible ratios) are squeezed in along the x-axis, the probability of producing any one ratio decreases.

cluded within the 95% limits are considered supportive of (failing to reject) the hypothesis of a 3:1 ratio. Any ratio in the remaining 5% area is considered unacceptable. (Other conventions also exist, such as rejection within the outer 10% or 1% limits; we consider these at the end of the chapter.) Thus, it is possible to see whether the experimental data support our hypothesis (in this case, the hypothesis of 3:1). One in twenty times (5%) we will make a **type I error:** We will reject a true hypothesis. (A **type II error** is failing to reject a false hypothesis.)

To determine whether to reject a hypothesis, we must derive a frequency distribution for each type of experiment. Mendel could have used the distribution shown in figure 4.1 for seed coat or seed color, as long as he was expecting a 3:1 ratio and had a similar sample size. What about independent assortment, which predicts a 9:3:3:1 ratio? A geneticist would have to calculate a new sampling distribution based on a 9:3:3:1 ratio and a particular sample size. Statisticians have devised shortcut methods by using standardized probability distributions. Many are in use, such as the *t*-distribution, binomial distribution, and chi-square distribution. Each is useful for particular kinds of data; geneticists usually use the chi-square distribution to test hypotheses regarding breeding data.

## Chi-Square

When sample subjects are distributed among discrete categories such as tall and dwarf plants, geneticists frequently use the **chi-square distribution** to evaluate data. The formula for converting categorical experimental data to a chi-square value is

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where $\chi$ is the Greek letter chi, $O$ is the observed number for a category, $E$ is the expected number for that category, and $\Sigma$ means to sum the calculations for all categories.

A chi-square ($\chi^2$) value of 0.60 is calculated in table 4.2 for Mendel's data on the basis of a 3:1 ratio. If Mendel had originally expected a 1:1 ratio, he would have calculated a chi-square of 244.45 (table 4.3). However, these $\chi^2$ values have little meaning in themselves: they are not probabilities. We can convert them to probabilities by determining where the chi-square value falls in relation to the area under the chi-square distribution curve. We usually use a chi-square table that contains probabilities that have already been calculated (table 4.4). Before we can use this table, however, we must define the concept of **degrees of freedom.**

Reexamination of the chi-square formula and tables 4.2 and 4.3 reveals that each category of data contributes to the total chi-square value, because chi-square is a summed value. We therefore expect the chi-square value to increase as the total number of categories increases. That is, the more categories involved, the larger the chi-square value, even if the sample fits relatively well against the hypothesized ratio. Hence, we need some way of keeping track of categories. We can do this with degrees of freedom, which is basically a count of independent categories. With Mendel's data, the total number of offspring is 1,064, of which 787 had tall stems. Therefore, the short-stem group had to consist of 277 plants (1,064 − 787) and isn't an independent category. For our purposes here, degrees of freedom equal the number of categories minus one. Thus, with two phenotypic categories, there is only one degree of freedom.

Table 4.4, the table of chi-square probabilities, is read as follows. Degrees of freedom appear in the left column. We are interested in the first row, where there is one degree of freedom. The numbers across the top of the table are the probabilities. We are interested in the next-to-the-last column, headed by the 0.05. We thus gain the following information from the table: *The probability is 0.05 of getting a chi-square value of 3.841 or larger by chance alone, given that the hypothesis is correct.* This statement formalizes the information in our discussion of frequency distributions. Hence, we are interested in how large a chi-square value will be found in the 5% unacceptable area of the curve. For Mendel's plant experiment, the **critical chi-square** (at $p = 0.05$, one degree of freedom) is 3.841. This is the value to which we compare the calculated $\chi^2$ values (0.60 and 244.45). Since the chi-square value for the 3:1 ratio is 0.60 (table 4.2), which is less than the critical value of 3.841, we do not reject the hypothesis of a 3:1 ratio. But since $\chi^2$ for the 1:1 ratio (table 4.3) is 244.45, which is greater than the critical value, we reject the hypothesis of a 1:1 ratio. Notice that once we did the chi-square test for the 3:1 ratio and failed to reject the hypothesis, no other statistical tests were needed: Mendel's data are consistent with a 3:1 ratio.

A word of warning when using the chi-square: If the expected number in any category is less than five, the conclusions are not reliable. In that case, you can repeat the experiment to obtain a larger sample size, or you can

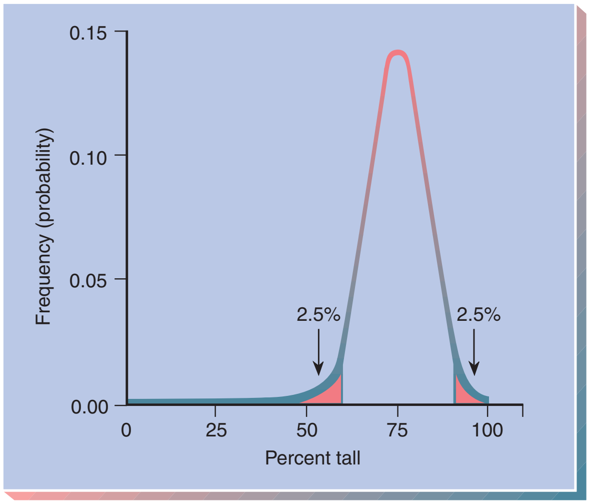


**Figure 4.2** Sampling distribution of figure 4.1, $n = 40$. By convention, 5% of the area is marked off (2.5% at each end).

combine categories. Note also that chi-square tests are always done on whole numbers, not on ratios or percentages.

## Failing to Reject Hypotheses

Hypothesis testing, in general, involves testing the assumption that there is no difference between the observed and the expected samples. Therefore, the hypothesis against which the data are tested is referred to as the **null hypothesis.** If the null hypothesis is not rejected, then we say that the data are consistent with it, not that the hypothesis has been proved. (As previously discussed, it is always possible we are not rejecting a false hypothesis or are rejecting the true one.) If, however, the hypothesis is rejected, as we rejected a 1:1 ratio for Mendel's data, we fail to reject the alternative hypothesis: that there is a difference between the observed and the expected values. We may then retest the data against some other hypothesis. (We don't say "accept the hypothesis" but rather "fail to reject the hypothesis," because supportive numbers could arise for many reasons. Our failure to reject is tentative acceptance of a hypothesis. However, we are on stronger ground when we reject a hypothesis.)

The use of the 0.05 probability level as a cutoff for rejecting a hypothesis is a convention called the **level of significance.** When a hypothesis is rejected at that level, statisticians say that the data depart *significantly* from the expected ratio. Other levels of significance are also

**Table 4.2** Chi-Square Analysis of One of Mendel's Experiments, Assuming a 3:1 Ratio

| | Tall Plants | Dwarf Plants | Total |
|---|---|---|---|
| Observed numbers ($O$) | 787 | 277 | 1,064 |
| Expected ratio | 3/4 | 1/4 | |
| Expected numbers ($E$) | 798 | 266 | 1,064 |
| $O - E$ | −11 | 11 | |
| $(O - E)^2$ | 121 | 121 | |
| $(O - E)^2/E$ | 0.15 | 0.45 | $0.60 = \chi^2$ |

**Table 4.3** Chi-Square Analysis of One of Mendel's Experiments, Assuming a 1:1 Ratio

| | Tall Plants | Dwarf Plants | Total |
|---|---|---|---|
| Observed numbers ($O$) | 787 | 277 | 1,064 |
| Expected ratio | 1/2 | 1/2 | |
| Expected numbers ($E$) | 532 | 532 | 1,064 |
| $O - E$ | 255 | −255 | |
| $(O - E)^2$ | 65,025 | 65,025 | |
| $(O - E)^2/E$ | 122.23 | 122.23 | $244.45 = \chi^2$ |

**Table 4.4** Chi-Square Values

| Degrees of Freedom | Probabilities | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0.99 | 0.95 | 0.80 | 0.50 | 0.20 | 0.05 | 0.01 |
| 1 | 0.000 | 0.004 | 0.064 | 0.455 | 1.642 | 3.841 | 6.635 |
| 2 | 0.020 | 0.103 | 0.446 | 1.386 | 3.219 | 5.991 | 9.210 |
| 3 | 0.115 | 0.352 | 1.005 | 2.366 | 4.642 | 7.815 | 11.345 |
| 4 | 0.297 | 0.711 | 1.649 | 3.357 | 5.989 | 9.488 | 13.277 |
| 5 | 0.554 | 1.145 | 2.343 | 4.351 | 7.289 | 11.070 | 15.086 |
| 6 | 0.872 | 1.635 | 3.070 | 5.348 | 8.558 | 12.592 | 16.812 |
| 7 | 1.239 | 2.167 | 3.822 | 6.346 | 9.803 | 14.067 | 18.475 |
| 8 | 1.646 | 2.733 | 4.594 | 7.344 | 11.030 | 15.507 | 20.090 |
| 9 | 2.088 | 3.325 | 5.380 | 8.343 | 12.242 | 16.919 | 21.666 |
| 10 | 2.558 | 3.940 | 6.179 | 9.342 | 13.442 | 18.307 | 23.209 |
| 15 | 5.229 | 7.261 | 10.307 | 14.339 | 19.311 | 24.996 | 30.578 |
| 20 | 8.260 | 10.851 | 14.578 | 19.337 | 25.038 | 31.410 | 37.566 |
| 25 | 11.524 | 14.611 | 18.940 | 24.337 | 30.675 | 37.652 | 44.314 |
| 30 | 14.953 | 18.493 | 23.364 | 29.336 | 36.250 | 43.773 | 50.892 |

used, such as 0.01. If a calculated chi-square is greater than the critical value in the table at the 0.01 level, we say that the data depart in a *highly significant* manner from the null hypothesis. Since the chi-square value at the 0.01 level is larger than the value at the 0.05 level, it is more difficult to reject a hypothesis at this level and hence more convincing when it is rejected. Other levels of rejection are also set. In clinical trials of medication, for example, experimenters attempt to make it very easy to reject the null hypothesis: a level of significance of 0.10 or higher is set. The rationale is that it is not desirable to discard a drug or treatment that may well be beneficial. Since the null hypothesis states that the drug has no effect—that is, the control and drug groups show the same response—clinicians would rather be overly conservative. Not rejecting the hypothesis means concluding that the drug has no effect. Rejecting the hypothesis means that the drug has some effect and should be tested further. It is much better to have to retest some drugs that are actually worthless than to discard drugs that have potential value.

## SUMMARY

**STUDY OBJECTIVE 1:** To understand the rules of probability and how they apply to genetics 71–74

We have examined the rules of probability theory relevant to genetic experiments. Probability theory allows us to predict the outcomes of experiments. The probability ($P$) of independent events occurring is calculated by multiplying their separate probabilities. The probability of mutually exclusive events occurring is calculated by adding their individual probabilities. And the probability of unordered events is defined by the polynomial expansion $(p + q + r + \ldots)^n$:

$$P = \frac{n!}{s!t!u!\ldots} p^s q^t r^u \ldots$$

**STUDY OBJECTIVE 2:** To understand the use of the chi-square statistical test in genetics 74–78

To assess whether data gathered during an experiment actually support a particular hypothesis, it is necessary to determine what the probability is of getting a particular data set when the null hypothesis is correct. One way to do this is through the chi-square test:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

This test gives us a method of quantifying the confidence we can place in the results obtained from typical genetic experiments. The rules of probability and statistics allow us to devise hypotheses about inheritance and to test these hypotheses with experimental data.

## SOLVED PROBLEMS

**PROBLEM 1:** Mendel self-fertilized a dihybrid plant that had round, yellow peas. In the offspring generation: What is the probability that a pea picked at random will be round and yellow? What is the probability that five peas picked at random will be round and yellow? What is the probability that of five peas picked at random, four will be round and yellow, and one will be wrinkled and green?

***Answer:*** The offspring peas will be round and yellow, round and green, wrinkled and yellow, and wrinkled and green in a ratio of 9:3:3:1. Thus, the probabilities that a pea picked at random will be one of these four categories are 9/16, 3/16, 3/16, and 1/16, respectively. Thus, the probability that a pea picked at random will be round and yellow is 9/16, or 0.563. The probability of picking five of these peas in a row is $(9/16)^5$ or 0.056. The probability that of five peas picked at random, four will be round and yellow, and one will be wrinkled and green, is (substituting into the binomial equation): $(5!/4!1!)(9/16)^4 (1/16)^1 = 5(9^4)/(16^5) = 5(0.006) = 0.031$.

**PROBLEM 2:** On a chicken farm, walnut-combed fowl that were crossed with each other produced the following offspring: walnut-combed, 87; rose-combed, 31; pea-combed, 30; and single-combed, 12. What hypothesis might you have (based on chapter 2) about the control of comb shape in fowl? Do the data support that hypothesis?

***Answer:*** The numbers 87, 31, 30, and 12 are very similar to 90, 30, 30, and 10, which would be a perfect fit to a 9:3:3:1 ratio. We might expect that ratio, having previously learned something about how comb type is inherited in fowl. Thus, we hypothesize that inheritance of comb type is by two loci, and that dominant alleles at both result in walnut combs, a dominant allele at one locus and recessives at the other result in rose or pea combs, and the recessive alleles at both loci result in a single comb. The results of the cross of dihybrids should produce fowl with the four comb types in a 9:3:3:1 ratio of walnut-, rose-, pea-, and single-combed fowl, respectively. Therefore, our observed numbers are 87, 31, 30, and 10 (sum = 160). Our expected ratio is 9:3:3:1, or 90, 30, 30, and 10 fowl, which are 9/16, 3/16, 3/16, and 1/16, respectively, of the sum of 160. We therefore set up the following chi-square table:

| | Comb Type | | | | |
|---|---|---|---|---|---|
| | **Walnut** | **Rose** | **Pea** | **Single** | **Total** |
| Observed Numbers (*O*) | 87 | 31 | 30 | 12 | 160 |
| Expected Ratio | 9/16 | 3/16 | 3/16 | 1/16 | |
| Expected Numbers (*E*) | 90 | 30 | 30 | 10 | 160 |
| $O - E$ | −3 | 1 | 0 | 2 | |
| $(O - E)^2$ | 9 | 1 | 0 | 4 | |
| $(O - E)^2/E$ | 0.1 | 0.033 | 0 | 0.4 | $0.533 = \chi^2$ |

There are three degrees of freedom since there are four categories of combs. (4 − 1 = 3). The critical chi-square value with three degrees of freedom and probability of 0.05 = 7.815 (table 4.4). Since our calculated chi-square value (0.533) is less than this critical value, we cannot reject our hypothesis. In other words, our data are consistent with the hypothesis of a 9:3:3:1 phenotypic ratio, indicative of a two-locus genetic model with dominance at each locus.

## EXERCISES AND PROBLEMS*

### PROBABILITY

1. Assuming a 1:1 sex ratio, what is the probability that five children produced by the same parents will consist of
   - **a.** three daughters and two sons?
   - **b.** alternating sexes, starting with a son?
   - **c.** alternating sexes?
   - **d.** all daughters?
   - **e.** all the same sex?
   - **f.** at least four daughters?
   - **g.** a daughter as the eldest child and a son as the youngest?

   (*See also* USE OF RULES)
2. Phenylthiocarbamide (PTC) tasting is dominant (*T*) to nontasting (*t*). If a taster woman with a nontaster father produces children with a taster man, and the man previously had a nontaster daughter, what would be the probability that
   - **a.** their first child would be a nontaster?
   - **b.** their first child would be a nontaster girl?
   - **c.** if they had six children, they would have two nontaster sons, two nontaster daughters, and two taster sons?
   - **d.** their fourth child would be a taster daughter?

   (*See also* USE OF RULES)
3. Albinism is recessive; assume for this problem that blue eyes are also recessive (albinos have blue eyes). What is the probability that two brown-eyed persons, heterozygous for both traits, would produce (remembering epistasis)
   - **a.** five albino children?
   - **b.** five albino sons?
   - **c.** four blue-eyed daughters and a brown-eyed son?
   - **d.** two sons genotypically like their father and two daughters genotypically like their mother?

   (*See also* USE OF RULES)
4. On the average, about one child in every ten thousand live births in the United States has phenylketonuria (PKU). What is the probability that
   - **a.** the next child born in a Boston hospital will have PKU?
   - **b.** after that child with PKU is born, the next child born will have PKU?
   - **c.** two children born in a row will have PKU?
5. In fruit flies, the diploid chromosome number is eight.
   - **a.** What is the probability that a male gamete will contain only paternal centromeres or only maternal centromeres?

---

*Answers to selected exercises and problems are on page A-4.

**b.** What is the probability that a zygote will contain only centromeres from male grandparents? (Disregard the problems that the sex chromosomes may introduce.)

**6.** How many seeds should Mendel have tested to determine with complete certainty that a plant with a dominant phenotype was heterozygous? With 99% certainty? With 95% certainty? With "pretty reliable" certainty?

**7.** What chance do a man and a woman have of producing one son and one daughter?

**8.** PKU and albinism are two autosomal recessive disorders, unlinked in human beings. If two people, each heterozygous for both traits, produce a child, what is the chance of their having a child with

**a.** PKU?
**b.** either PKU or albinism?
**c.** both traits?

**9.** In human beings, the absence of molars is inherited as a dominant trait. If two heterozygotes have four children, what is the probability that

**a.** all will have no molars?
**b.** three will have no molars and one will have molars?
**c.** the first two will have molars and the second two will have no molars?

(*See also* USE OF RULES)

**10.** Galactosemia is inherited as a recessive trait. If two normal heterozygotes produce children, what is the chance that

**a.** one of four children will be affected?
**b.** three children will be born in this order: normal boy, affected girl, affected boy?

(*See also* USE OF RULES)

**11.** A normal man (A) whose grandfather had galactosemia and a normal woman (B) whose mother was galactosemic want to produce a child. What is the probability that their first child will be galactosemic?

**12.** A city had nine hundred deaths during the year, and of these, three hundred were from cancer and two hundred from heart disease. What is the probability that the next death will be from

**a.** cancer?
**b.** either cancer or heart disease?

**13.** A plant that has the genotype *AA bb cc DD EE* is mated with one that is *aa BB CC dd ee.* $F_1$ individuals are selfed. What is the chance of getting an $F_2$ plant whose genotype exactly matches the genotype of one of the parents?

## USE OF RULES

**14.** The ability to taste phenylthiocarbamide is dominant in human beings. If a heterozygous taster mates with a nontaster, what is the probability that of their five children, only one will be a taster?

**15.** In mice, coat color is determined by two independent genes, A and C, as indicated here: $A\text{-}C\text{-}$, agouti; $aaC\text{-}$, black; $A\text{-}c^ac^a$ and $aac^ac^a$, albino. If the following two mice are crossed $AaCc^a \times Aac^ac^a$, what is the probability that among the first six offspring, two will be agouti, two will be black, and two will be albino?

## STATISTICS

**16.** The following data are from Mendel's original experiments. Suggest a hypothesis for each set and test this hypothesis with the chi-square test. Do you reach different conclusions with different levels of significance?

**a.** Self-fertilization of round-seeded hybrids produced 5,474 round seeds and 1,850 wrinkled ones.
**b.** One particular plant from *a* yielded 45 round seeds and 12 wrinkled ones.
**c.** Of the 565 plants raised from $F_2$ round-seeded plants, 372, when self-fertilized, gave both round and wrinkled seeds in a 3:1 proportion, whereas 193 yielded only round seeds.
**d.** A violet-flowered, long-stemmed plant was crossed with a white-flowered, short-stemmed plant, producing the following offspring:

47 violet, long-stemmed plants
40 white, long-stemmed plants
38 violet, short-stemmed plants
41 white, short-stemmed plants

**17.** Mendel self-fertilized pea plants with round and yellow peas. In the next generation he recovered the following numbers of peas:

315 round and yellow peas
108 round and green peas
101 wrinkled and yellow peas
32 wrinkled and green peas

What is your hypothesis about the genetic control of the phenotype? Do the data support this hypothesis?

**18.** Two agouti mice are crossed, and over a period of a year they produce 48 offspring with the following phenotypes:

28 agouti mice
7 black mice
13 albino mice

What is your hypothesis about the genetic control of coat color in these mice? Do the data support that hypothesis?

**19.** Two curly-winged flies, when mated, produce sixty-one curly and thirty-five straight-winged progeny. Use a chi-square test to determine whether these numbers fit a 3:1 ratio.

**20.** A short-winged, dark-bodied fly is crossed with a long-winged, tan-bodied fly. All the $F_1$ progeny are long-winged and tan-bodied. $F_1$ flies are crossed among themselves to yield 84 long-winged, tan-bodied flies; 27 long-winged dark-bodied flies; 35 short-winged, tan-bodied flies; and 14 short-winged, dark-bodied flies.

**a.** What ratio do you expect in the progeny?

**b.** Use the chi-square test to evaluate your hypothesis. Is the observed ratio within the expected range?

## CRITICAL THINKING QUESTIONS

**1.** A friend shows you three closed boxes, one of which contains a prize, and asks you to choose one. Your friend then opens one of the two remaining boxes, a box she knows is empty. At that point, she gives you the opportunity to change your choice to the last remaining box. Should you?

**2.** If all couples wanted at least one child of each sex, approximately what would the average family size be?

*Suggested Readings for chapter 4 are on page B-2.*

# 5

# SEX DETERMINATION, SEX LINKAGE, AND PEDIGREE ANALYSIS

## STUDY OBJECTIVES

1. To analyze the causes of sex determination in various organisms 83
2. To understand methods of dosage compensation 90
3. To analyze the inheritance patterns of traits that loci on the sex chromosomes control 95
4. To use pedigrees to infer inheritance patterns 97

## STUDY OUTLINE

**Sex Determination** 83
  Patterns 83
  Sex Chromosomes 83
  Sex Determination in Flowering Plants 87
**Dosage Compensation** 90
  Proof of the Lyon Hypothesis 90
  Dosage Compensation for *Drosophila* 94
**Sex Linkage** 95
  X Linkage in *Drosophila* 95
  Nonreciprocity 96
  Sex-Limited and Sex-Influenced Traits 96
**Pedigree Analysis** 97
  Penetrance and Expressivity 97
  Family Tree 98
  Dominant Inheritance 99
  Recessive Inheritance 99
  Sex-Linked Inheritance 100
*Summary* 103
*Solved Problems* 103
*Exercises and Problems* 104
*Critical Thinking Questions* 108
**Box 5.1 Why Sex and Why Y?** 88
**Box 5.2 Electrophoresis** 92

Three generations of a family.

(© Frank Siteman/Tony Stone Images.)

We ended chapter 3 with a discussion of the chromosomal theory of heredity, stated lucidly in 1903 by Walter Sutton, that genes are located on chromosomes. In 1910, T. H. Morgan, a 1933 Nobel laureate, published a paper on the inheritance of white eyes in fruit flies. The mode of inheritance for this trait, discussed later in this chapter, led inevitably to the conclusion that the locus for this gene is on a chromosome that determines the sex of the flies: when a white-eyed male was mated with a red-eyed female, half of the $F_2$ sons were white-eyed and half were red-eyed; all $F_2$ daughters were red-eyed. Not only was this the first evidence that localized a particular gene to a particular chromosome, but this study also laid the foundation for our understanding of the genetic control of sex determination.

## SEX DETERMINATION

### Patterns

At the outset, we should note that the sex of an organism usually depends on a very complicated series of developmental changes under genetic and hormonal control. However, often one or a few genes can determine which pathway of development an organism takes. Those switch genes are located on the **sex chromosomes,** a heteromorphic pair of chromosomes, when those chromosomes exist.

However, sex chromosomes are not the only determinants of an organism's sex. The ploidy of an individual, as in many hymenoptera (bees, ants, wasps), can determine sex; males are haploid and females are diploid. Allelic mechanisms may determine sex by a single allele or multiple alleles not associated with heteromorphic chromosomes; even environmental factors may control sex. For example, temperature determines the sex of some geckos, and the sex of some marine worms and gastropods depends on the substrate on which they land. In this chapter, however, we concentrate on chromosomal sex-determining mechanisms.

### Sex Chromosomes

Basically, four types of chromosomal sex-determining mechanisms exist: the XY, ZW, X0, and compound chromosomal mechanisms. In the XY case, as in human beings or fruit flies, the females have a homomorphic pair of chromosomes (XX) and males are heteromorphic (XY). In the ZW case, males are homomorphic (ZZ), and females are heteromorphic (ZW). (XY and ZW are chromosome notations and imply nothing about the sizes or shapes of these chromosomes.) In the X0 case, the organism has only one sex chromosome, as in some grasshoppers and beetles; females are usually XX and males X0. And in the compound chromosome case, several X and Y chromosomes combine to determine sex, as in bedbugs and some beetles. We need to emphasize that the chromosomes themselves do not determine sex, but the genes they carry do. In general, the genotype determines the type of gonad, which then determines the phenotype of the organism through male or female hormonal production.

### The XY System

The XY situation occurs in human beings, in which females have forty-six chromosomes arranged in twenty-three homologous, homomorphic pairs. Males, with the same number of chromosomes, have twenty-two homomorphic pairs and one heteromorphic pair, the XY pair (fig. 5.1). During meiosis, females produce gametes that contain only the X chromosome, whereas males produce two kinds of gametes, X- and Y-bearing (fig. 5.2). For this reason, females are referred to as **homogametic** and males as **heterogametic.** As you can see from figure 5.2, in people, fertilization has an equal chance of producing either male or female offspring. In *Drosophila,* the system is the same, but the Y chromosome is almost 20% larger than the X chromosome (fig. 5.3).

Since both human and *Drosophila* females normally have two X chromosomes, and males have an X and a Y chromosome, it seems impossible to know whether maleness is determined by the presence of a Y chromosome or the absence of a second X chromosome. One way to resolve this problem would be to isolate individuals with odd numbers of chromosomes. In chapter 8, we examine the causes and outcomes of anomalous chromosome numbers. Here, we consider two facts from that chapter. First, in rare instances, individuals form, although they are not necessarily viable, with extra sets of chromosomes. These individuals are referred to as **polyploids** (*triploids* with $3n$, *tetraploids* with $4n$, etc.). Second, also infrequently, individuals form that have more or fewer than the normal number of any one chromosome. These **aneuploids** usually come about when a pair of chromosomes fails to separate properly during meiosis, an occurrence called **nondisjunction.** The existence of polyploid and aneuploid individuals makes it possible to test whether the Y chromosome is male determining. For example, a person or a fruit fly that has all the proper nonsex chromosomes, or **autosomes** (forty-four in human beings, six in *Drosophila*), but only a single X without a Y would answer our question. If the Y were absolutely male determining, then this X0 individual should be female. However, if the sex-determining mechanism is a result of the number of X chromosomes, this individual should be a male. As it turns out, an X0 individual is a *Drosophila* male and a human female.

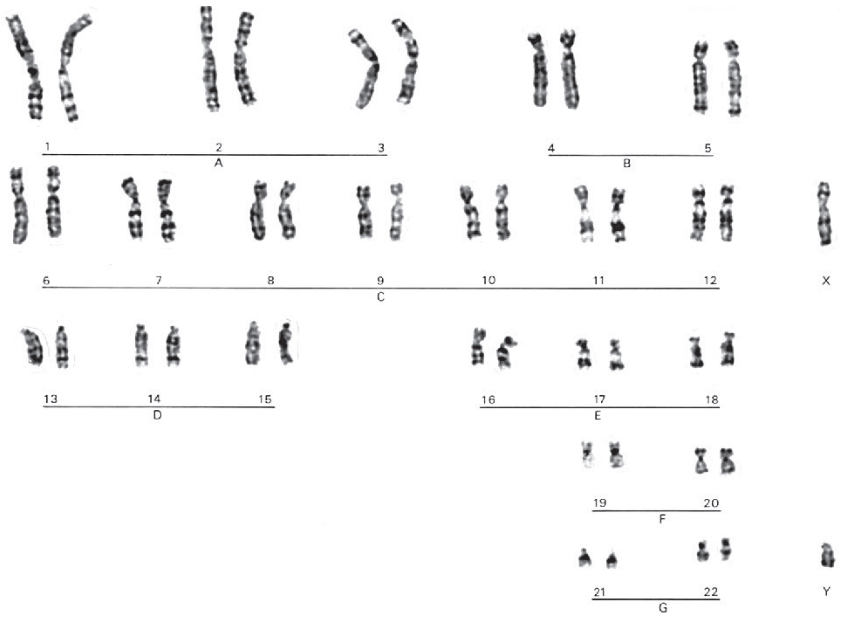
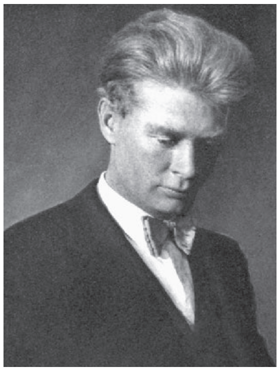


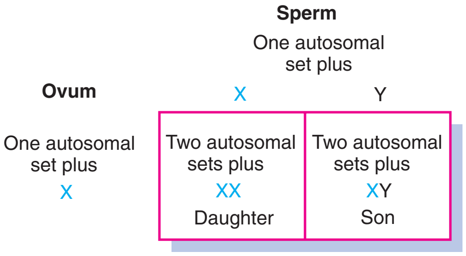
**Figure 5.1** Human male karyotype. Note the X and Y chromosomes. A female would have a second X chromosome in place of the Y. (Reproduced courtesy of Dr. Thomas G. Brewster, Foundation for Blood Research, Scarborough, Maine.)

### *Genic Balance in* Drosophila

When geneticist Calvin Bridges, working with *Drosophila,* crossed a triploid (3*n*) female with a normal male, he observed many combinations of autosomes and sex chromosomes in the offspring. From his results, Bridges suggested in 1921 that sex in *Drosophila* is determined by the balance between (ratio of) autosomal alleles that favor maleness and alleles on the X chromosomes that favor femaleness. He calculated a ratio of X chromosomes to autosomal sets to see if this ratio would predict the sex of a fly. An **autosomal set** (A) in *Drosophila* consists of one chromosome from each autosomal pair, or three chromosomes. (An autosomal set in human beings consists of twenty-two chromosomes.) Table 5.1, which presents his results, shows that Bridges's **genic balance**



Calvin B. Bridges (1889–1938). (From *Genetics* 25 (1940): frontispiece. Courtesy of the Genetics Society of America.)



| Ovum | Sperm: One autosomal set plus X | Sperm: One autosomal set plus Y |
|---|---|---|
| One autosomal set plus X | Two autosomal sets plus XX<br>Daughter | Two autosomal sets plus XY<br>Son |

**Figure 5.2** Segregation of human sex chromosomes during meiosis, with subsequent zygote formation.

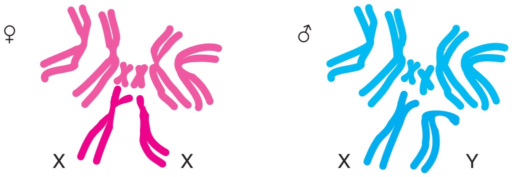


**Figure 5.3** Chromosomes of *Drosophila melanogaster.*

**Table 5.1** Data Supporting Bridges's Theory of Sex Determination by Genic Balance in *Drosophila*

| Number of X Chromosomes | Number of Autosomal Sets (A) | Total Number of Chromosomes | $\frac{X}{A}$ Ratio | Sex |
|---|---|---|---|---|
| 3 | 2 | 9 | 1.50 | Metafemale |
| 4 | 3 | 13 | 1.33 | Female |
| 4 | 4 | 16 | 1.00 | Female |
| 3 | 3 | 12 | 1.00 | Female |
| 2 | 2 | 8 | 1.00 | Female |
| 1 | 1 | 4 | 1.00 | Female |
| 2 | 3 | 11 | 0.67 | Intersex |
| 1 | 2 | 7 | 0.50 | Male |
| 1 | 3 | 10 | 0.33 | Metamale |

**theory** of sex determination was essentially correct. When the X:A ratio is 1.00, as in a normal female, or greater than 1.00, the organism is a female. When this ratio is 0.50, as in a normal male, or less than 0.50, the organism is a male. At 0.67, the organism is an **intersex. Metamales** (X/A = 0.33) and **metafemales** (X/A = 1.50) are usually very weak and sterile. The metafemales usually do not even emerge from their pupal cases.

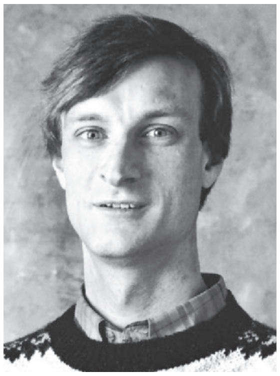
A **sex-switch** gene has been discovered that directs female development. This gene, **Sex-lethal** (*Sxl*), is located on the X chromosome. (It was originally called *female-lethal* because mutations of this gene killed female embryos.) Apparently, *Sxl* has two states of activity. When it is "on," it directs female development; when it is "off," maleness ensues. Other genes located on the X chromosome and the autosomes regulate this sex-switch gene. Genes on the X chromosome that act to regulate *Sxl* into the on state (female development) are called **numerator elements** because they act on the numerator of the X/A genic balance equation. Genes on the autosomes that act to regulate *Sxl* into the off state (male development) are called **denominator elements.** Geneticists have discovered four numerator elements—genes named *sisterless-a, sisterless-b, sisterless-c,* and *runt*. *Sxl* "counts" the number of X chromosomes; it turns on when two are present. It counts by measuring the level of the numerator genes' protein product. If the level is high, *Sxl* turns on, and the organism develops as a female. If the level is relatively low, *Sxl* does not turn on, and development proceeds as a male.

### Sex Determination in Human Beings

Since the X0 genotype in human beings is a female (having Turner syndrome), it seems reasonable to conclude that the Y chromosome is male determining in human beings. The fact that persons with Klinefelter syndrome (XXY, XXXY, XXXXY) are all male, and XXX, XXXX, and other multiple-X karyotypes are all female, verifies this idea. (More details on these anomalies are presented in chapter 8.) For a long time, researchers have sought a single gene, a **testis-determining factor (TDF),** located on the Y chromosome that acts as a sex switch to initiate male development. Human embryologists had discovered that during the first month of embryonic development, the gonads that develop are neither testes nor ovaries, but instead are indeterminate. At about six or seven weeks of development, the indeterminate gonads become either ovaries or testes.

In the 1950s, Ernst Eichwald found that males had a protein on their cell surfaces not found in females; he discovered that female mice rejected skin grafts from genetically identical brothers, whereas the brothers accepted grafts from sisters. This implies that an antigen exists on the surface of male cells that is not found on female cells. This protein was called the *histocompatibility Y antigen* **(H-Y antigen).** The gene for this protein was found on the Y chromosome, near the centromere. At first, scientists believed it to be the sex switch: if the gene were present, the gonads would begin development as testes. Further male development, as in male secondary sexual characteristics, came about through the testosterone the functional testes produced. If the gene were absent, the gonads would develop into ovaries. Recently, however, by studying "sex-reversed" individuals, biologists refuted this theory.

Sex-reversed individuals are XX males or XY females. David Page, at the Whitehead Institute for Biomedical Research, found twenty XX males who had a small piece of the short arm of the Y chromosome attached to one of their X chromosomes. He found six XY females in whom the Y chromosome was missing the same small piece at the end of its short arm. This region, which did not contain the *HYA* gene, must carry the testis-determining factor. The first candidate gene from this region believed to code for the testis-determining factor was named the



David Page (1956– ).
(Courtesy of Dr. David Page.)

***ZFY* gene,** for zinc finger on the Y chromosome. Zinc fingers are protein configurations known to interact with DNA (discussed in detail in chapter 16). Thus, researchers believed that the *ZFY* gene, coding for the testis-determining factor, worked by directly interacting with DNA. (Later in the book we look at the way regulatory genes, whose proteins interact with DNA, work.) However, men who lack the *ZFY* gene have been found, suggesting that the testis-determining factor is very close to, but not, the *ZFY* gene. From work in mice, it has been suggested that the *ZFY* gene controls the initiation of sperm cell development, but not maleness.

In 1991, Robin Lovell-Badge and Peter Goodfellow and their colleagues in England isolated a gene called **Sex-determining region Y (*SRY*)**—*Sry* in mice—adjacent to the *ZFY* gene. *Sry* has been positively identified as the testis-determining factor because, when injected into normal (XX) female mice, it caused them to develop as males (fig. 5.4). Although these XX males are sterile, they appear as normal males in every other way. (We discuss in chapter 13 how scientists introduce new genes into an organism.) Note also that the mouse and human systems are very similar genetically, and the homologous genes have been isolated from both. However, at present, the human *SRY* gene does not convert XX female mice into males. Like the *ZFY* gene product, Sry protein (the protein the *SRY* gene produces) also binds to DNA.

Robin Lovell-Badge (1953– ).
(Courtesy of Robin Lovell-Badge.)

Peter Goodfellow (1951– ).
(Courtesy of Peter Goodfellow.)

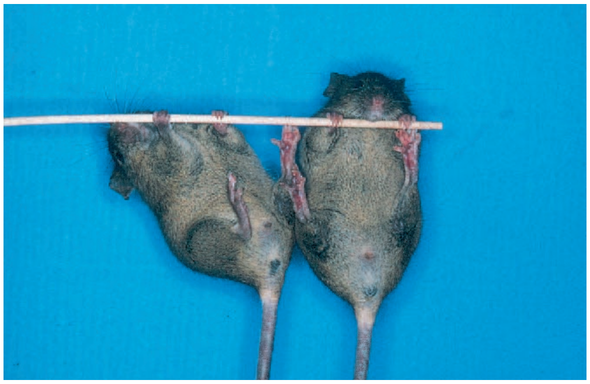


**Figure 5.4** Normal male mouse (*left*) and female littermate given the *Sry* gene (*right*). Both mice are indistinguishably male.
(Courtesy of Robin Lovell-Badge.)

The Sry protein appears to bind to at least two genes. One, the p450 aromatase gene, has a protein product that converts the male hormone testosterone to the female hormone estradiol; the Sry protein inhibits production of p450 aromatase. The second gene the Sry protein affects is the gene for the Müllerian-inhibiting substance, which induces testicular development and the digression of female reproductive ducts; the Sry protein enhances this gene's activity. Thus, the Sry protein points an indifferent embryo toward maleness and the maintenance of testosterone production. The sex switch initiates a developmental sequence involving numerous genes. Eva Eicher and Linda Washburn have developed a model in which two pathways of coordinated gene action help determine sex, one pathway for each sex. The first gene in the ovary-determining pathway is termed *ovary determining* (*Od*). The first gene in the testis-determining pathway must function before the *Od* gene begins, in order to allow XY individuals to develop as males. Once the steps of a pathway are initiated, the other pathway is inhibited (fig. 5.5).

### *Other Chromosomal Systems*

The X0 system, sometimes referred to as an X0-XX system, occurs in many species of insects. It functions just as the XY chromosomal mechanism does, except that instead of a Y chromosome, the heterogametic sex (male) has only one X chromosome. Males produce gametes that contain either an X chromosome or no sex chromosome, whereas all the gametes from a female contain the X chromosome. The result of this arrangement is that females have an even number of chromosomes (all in homomorphic pairs) and males have an odd number of chromosomes.

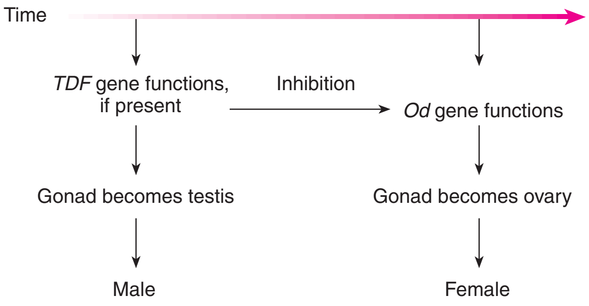


**Figure 5.5** A model for the initiation of gonad determination in mammals.

The ZW system is identical to the XY system except that males are homogametic and females are heterogametic. This situation occurs in birds, some fishes, and moths.

Compound chromosomal systems tend to be complex. For example, *Ascaris incurva,* a nematode, has eight X chromosomes and one Y. The species has twenty-six autosomes. Males have thirty-five chromosomes (26A + 8X + Y), and females have forty-two chromosomes (26A + 16X). During meiosis, the X chromosomes unite end to end and so behave as one unit.

### *The Y Chromosome*

In both human beings and fruit flies, the Y chromosome has very few functioning genes. In human beings, two homologous regions exist, one at either end of the X and Y chromosomes, allowing the chromosomes to pair during meiosis. These regions are termed **pseudoautosomal.** Mapping the Y chromosome (see chapters 6 and 13) has shown us the existence of about thirty-five genes (fig. 5.6). Other, nonfunctioning genes are present, too, remnants of a time in the evolutionary past when those genes were probably active (box 5.1). The *Drosophila* Y chromosome is known to carry genes for at least six fertility factors, two on the short arm (*ks-1* and *ks-2*) and four on the long arm (*kl-1, kl-2, kl-3,* and *kl-5*). The Y chromosome carries two other known genes: *bobbed,* which is a locus of ribosomal RNA genes (the nucleolar organizer), and *Suppressor of Stellate* or *Su(Ste),* a gene required for RNA splicing (see chapter 10). The fertility factors code for proteins needed during spermatogenesis. For example, *kl-5* codes for part of the dynein motor needed for sperm flagellar movement.

## Sex Determination in Flowering Plants

Flowering plant species (angiosperms) generally have three kinds of flowers: hermaphroditic, male, and female. Hermaphroditic flowers have both male and female parts. The male parts are the anthers and filaments, making up the *stamen,* and the female parts are the stigma, style, and ovary, making up the *pistil* (see fig. 2.2). Ninety percent of angiosperms have hermaphroditic flowers. Of the 10% of the species that have unisexual flowers, some are *monoecious* (Greek, one house), bearing both male and female flowers on the same plant (e.g., walnut); and some are *dioecious* (Greek, two houses), having plants with just male or just female flowers (e.g., date palm).

Within the group of plant species with unisexual flowers, sex-determining mechanisms vary. Some species have a single locus determining sex, some have two or more loci involved in sex determination, and some have X and Y chromosomes. In most of the species with X and Y chromosomes, the sex chromosomes are indistinguishable. Among these species, most have heterogametic males, although in some species, such as the strawberry, females are heterogametic. In the very few species that have distinguishable X and Y chromosomes—only thirteen are known—two sex-determination mechanisms are found. One is similar to the system in mammals, in which the Y chromosome has a gene or genes present

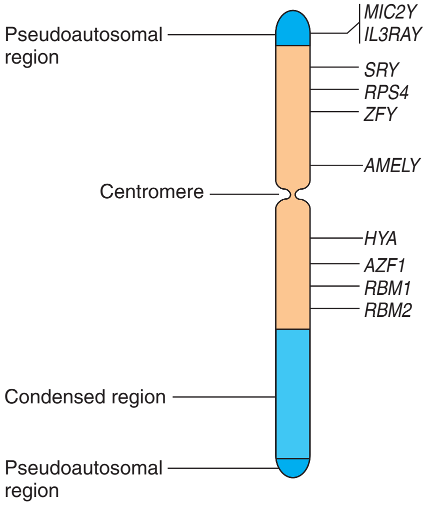


**Figure 5.6** The human Y chromosome. In addition to the genes shown, the Y chromosome carries other genes, homologous to X chromosome genes, that do not function because of accumulated mutations. Some of these are in multiple copies. Note the two pseudoautosomal regions that allow synapsis between the Y and X chromosomes. The gene symbols shown include *MIC2Y,* T cell adhesion antigen; *IL3RAY,* interleukin-3 receptor; *RPS4,* a ribosomal protein; *AMELY,* amelogenin; *HYA,* histocompatibility Y antigen; *AZF1,* azoospermia factor 1 (mutants result in tailless sperm); and *RBM1, RBM2,* RNA binding proteins 1 and 2. (Adapted from Online Mendelian Inheritance in Man website. http://www3.ncbi.nlm.nih.gov/omim/. Reprinted with permission.)

## BOX 5.1

## Experimental Methods

### *Why Sex and Why Y?*

Evolutionary biologists have asked, Why does sex exist? A haploid, asexual way of life seems like a very efficient form of existence. Haploid fungi can produce thousands of haploid spores, each of which can grow into a new colony. What evolutionary benefit do organisms gain by developing diploidy and sexual processes? Although this may not seem like a serious question, evolutionary biologists look for compelling answers.

In chapter 21, we discuss evolutionary thinking in some detail. For the moment, accept that evolutionary biologists look for an adaptive advantage in most evolutionary outcomes. Thus they ask, What is better about the combining of gametes to produce a new generation of offspring? Why would a diploid organism take a random sample of its genome and combine it with a random sample of someone else's genome to produce offspring? Why not simply produce offspring by mitosis? If offspring are produced by mitosis, all of an individual's genes pass into the next generation with every offspring. Not only does just half the genome of an individual pass into the next generation with every offspring produced sexually, but that half is a random jumble of what might be a very highly adapted genome. In addition, males are doubly expensive to produce because males themselves do not produce offspring: males fertilize females who produce offspring. Thus, on the surface, evolutionary biologists need to find very strong reasons for an organism to turn to sexual reproduction when an individual might be at an advantage evolutionarily to reproduce asexually.

There have been numerous suggestions as to the advantage of sex, nicely summarized in a 1994 article by James Crow, of the University of Wisconsin, in *Developmental Genetics,* and more recently in a special section of the 25 September 1998 issue of *Science* magazine. We aren't really sure what the true evolutionary reasons for sex are, but at least three explanations seem reasonable to evolutionary biologists:

- *Adjusting to a changing environment.* Sexual reproduction allows for much more variation in organisms. A haploid, asexual organism collects variation over time by mutation. A sexual organism, on the other hand, can achieve a tremendous amount of variation by recombination and fertilization. Remember that a human being can produce potentially $2^{100,000}$ different gametes. In a changing environment, a sexually reproduced organism is much more likely than an asexual organism to produce offspring that will be adapted to the changes.
- *Combining beneficial mutations.* As mentioned, a haploid, asexual organism accrues mutations as they happen over time in a given individual. A sexual organism can combine beneficial mutations each generation by recombination and fertilization. Thus, sexually reproducing organisms can adapt at a much more rapid rate than asexual organisms.
- *Removing deleterious mutations.* Mutation is more likely to produce deleterious changes than beneficial ones. An asexual organism gathers more and more deleterious mutations as time goes by (a process referred to as *Muller's ratchet,* in honor of Nobel Prize-winning geneticist H. J. Muller and referring to a ratchet wheel that can only go forward). Sexually reproducing organisms can eliminate deleterious mutations each generation by forming recombined offspring that are relatively free of mutation.

Hence, this list provides three of the generally assumed advantages of sexual reproduction that offset its disadvantages.

Another subtle question about sexual reproduction that evolutionary biologists ask is, Why is there a Y chromosome? In other words, why do we have, in some species (e.g., people), a heteromorphic pair of chromosomes involved in sex determination, with one of the chromosomes having the gene for that sex and very few other loci? In people, the Y chromosome is basically a degenerate chromosome with very few loci. This morphological difference between the members of the sex chromosome pair is puzzling. After all, chromosome pairs that do not carry sex-determining loci do not tend to be morphologically heterogeneous. Consider the following possible scenario that Virginia Morell presented in the 14 January 1994 issue of *Science.*

In a particular species in the past—evolutionarily speaking—a sex-determining gene arises on a particular chromosome. One allele at this locus confers maleness on its bearer. The absence of this allele causes the carrier to be female. At this point, millions of years ago, the sex chromosomes are not morphologically heterogeneous: the X and Y chromosomes are identical. In time, however, the Y chromosome comes to carry a gene that is beneficial to the male but not the female. For example, there might be a gene with an allele for a colorful marking; this allele confers a reproductive advantage for the male but also confers a predatory risk on the bearer, whether male or female. Males have a reproductive advantage to outweigh the predation risk, whereas females have none. Thus, the allele is favored in males and selected against in females.

An evolutionary solution to this situation is to isolate the gene for this marking on the Y chromosome and keep it off the X chromosome so that males have it but females do not. This can take place if the two chromosomes do not recombine over most of their lengths. Assume then, that some mechanism evolves to prevent recombination of the X and Y chromosomes. Thereafter, the Y chromosome degenerates, losing most of its genes but retaining the sex-determining locus and the loci conferring an advantage on males but a disadvantage on females.

What evidence do we have that any of these links in this complex line of logic are true? To begin with, when we look at evolutionary lineages, we usually see a spectrum of species with sex chromosomes in all stages of differentiation. Evolutionary biologists generally accept the notion that the similar sex chromosomes are the original condition and the morphologically heterogeneous sex chromosomes are the more evolved condition. In addition, as reported in the same issue of *Science,* William Rice of the University of California at Santa Cruz has shown experimentally with fruit flies that if recombination is prevented between sex chromosomes, the Y chromosome degenerates; it loses the function of many loci that are also found on the X chromosome. Rice showed this with an ingenious set of experiments that successfully prevented a nascent Y chromosome from recombining with the X. The results confirmed the prediction that the Y chromosome degenerates (fig. 1).

More recently, in an October 1999 article in *Science,* Bruce Lahn and David Page, at the Massachusetts Institute of Technology, reported research findings indicating that degeneration of the human Y chromosome has taken place in four stages, starting as long as 320 million years ago in our mammalian ancestors. Using DNA sequence data and methods discussed in chapter 21, they showed that the 19 genes known from both the X and Y chromosomes are arranged as if the Y chromosome has undergone four rearrangements, each preventing further recombination of the X and Y. According to their calculations, this process began shortly after the mammals split from the birds, which themselves went on to evolve a ZW pair of sex chromosomes.

Clearly, much more work is needed to validate all the steps in this logical, evolutionary argument. However, at this point, enough empirical support exists to make the idea attractive to evolutionary biologists.

Although we have gotten a bit ahead of ourselves by talking about subtle evolutionary arguments before reaching that material in the book, it is a good idea to keep an evolutionary perspective on processes and structures. Presumably, evolution has shaped us and the biological world in which we live. If that is so, we should be able to figure out how evolution was working. That thinking should hold from the level of the molecule (e.g., enzymes and DNA) to that of the whole organism. Behind every process and structure should be a hint of the evolutionary pressures that caused that structure or process to evolve.

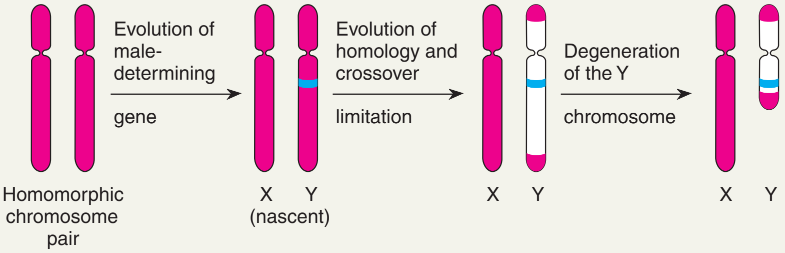

Evolution of male-determining gene
Evolution of homology and crossover limitation
Degeneration of the Y chromosome
Homomorphic chromosome pair
X
Y
(nascent)
X
Y
X
Y

**Figure 1** Evolution of a hypothetical Y chromosome. *Red* represents homologous regions, *blue* shows the male-determining gene, and *white* marks evolved areas of the Y chromosome that no longer recombine with the X chromosome.

that actively determine male-flowering plants. The other system is similar to that found in fruit flies, in which the X:A ratio determines sex.

In the mammalian-type system, the Y chromosome carries genes needed for the development of male flower parts while suppressing the development of female parts. An example of this is in the white campion *(Silene latifolia).* In the *Drosophila*-type system, found in the sorrel *(Rumex acetosa),* the ratios determine sex exactly as in the flies. That is, an X:A ratio of 0.5 or lower results in a male; a ratio of 1.0 or higher results in a female; and an intermediate ratio results in a plant with hermaphroditic flowers. It seems that all flowers have the potential to be hermaphroditic. That is, flower primordia for hermaphroditic, male, and female flowers look identical during early development. The simplest mechanism of sex determination would involve repressing the development of the female flower parts in male flowers and repressing the male flower parts in female flowers. Current research indicates that this repression of one component or another is probably involved in most flower sex determination and is under genetic and hormonal control. (We discuss further the genetic control of flower development in chapter 16.)

## DOSAGE COMPENSATION

In the XY chromosomal system of sex determination, males have only one X chromosome, whereas females have two. Thus, disregarding pseudoautosomal regions, males have half the number of X-linked alleles as females for genes that are not primarily related to gender. A question arises: How does the organism compensate for this dosage difference between the sexes, given the potential for serious abnormality? In general, an incorrect number of autosomes is usually highly deleterious to an organism (see chapter 8). In human beings and other mammals, the necessary **dosage compensation** is accomplished by the inactivation of one of the X chromosomes in females so that both males and females have only one functional X chromosome per cell.

In 1949, M. Barr and E. Bertram first observed a condensed body in the nucleus that was not the nucleolus. Noting that normal female cats show a single condensed body, while males show none, these researchers referred to the body as sex chromatin, since known as a **Barr body** (fig 5.7). Mary Lyon then suggested that this Barr body represented an inactive X chromosome, which in females becomes tightly coiled into heterochromatin, a condensed, and therefore visible, form of chromatin.

Various lines of evidence support the **Lyon hypothesis** that only one X chromosome is active in any cell. First, XXY males have a Barr body, whereas X0 females have none. Second, persons with abnormal numbers of X chromosomes have one fewer Barr body than they have X chromosomes per cell: XXX females have two Barr bodies and XXXX females have three.

Mary F. Lyon (1925– ).
(Courtesy of Dr. Mary F. Lyon.)

### Proof of the Lyon Hypothesis

Direct proof of the Lyon hypothesis came when cytologists identified the Barr body in normal females as an X chromosome. Genetic evidence also supports the Lyon hypothesis: Females heterozygous for a locus on the X chromosome show a unique pattern of phenotypic expression. We now know that in human females, an X chromosome is inactivated in each cell on about the twelfth day of embryonic life; we also know that the inactivated X is randomly determined in a given cell. From that point on, the same X remains a Barr body for future cell generations. Thus, heterozygous females show **mosaicism** at the cellular level for X-linked traits. Instead of being typically heterozygous, they express only one or the other of the X chromosomal alleles in each cell.

Glucose-6-phosphate dehydrogenase (G-6-PD) is an enzyme that a locus on the X chromosome controls. The

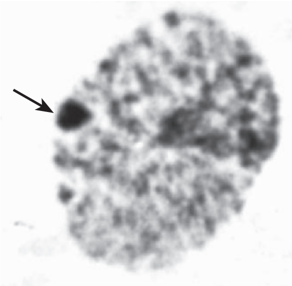


**Figure 5.7** Barr body (*arrow*) in the nucleus of a cheek mucosal cell of a normal woman. This visible mass of heterochromatin is an inactivated X chromosome. (Thomas G. Brewster and Park S. Gerald, "Chromosome disorders associated with mental retardation," *Pediatric Annals,* 7, no. 2, 1978. Reproduced courtesy of Dr. Thomas G. Brewster, Foundation for Blood Research, Scarborough, Maine.)

enzyme occurs in several different allelic forms that differ by single amino acids. Thus, both forms (A and B) will dehydrogenate glucose-6-phosphate—both are fully functional enzymes—but because they differ by an amino acid, they can be distinguished by their rate of migration in an electrical field (one form moves faster than another). This electrical separation, termed **electrophoresis,** is carried out by placing samples of the enzymes in a supporting gel, usually starch, polyacrylamide, agarose, or cellulose acetate (fig. 5.8 and box 5.2). After the electric current is applied for several hours, the enzymes move in the gel as bands, revealing the distance each enzyme traveled. Since blood serum is a conglomerate of proteins from many cells, the serum of a female heterozygote (fig. 5.8, lane 3) has both A and B forms (bands), whereas any single cell (lanes 4–10) has only one or the other. Since the gene for glucose-6-phosphate dehydrogenase is carried on the X chromosome, this electrophoretic display indicates that only one X is active in any particular cell.

Another aspect of the glucose-6-phosphate dehydrogenase system provides further proof of the Lyon hypothesis. If a cell has both alleles functioning, both A and B proteins should be present. Since the functioning glucose-6-phosphate dehydrogenase enzyme is a dimer (made up of two protein subunits), 50% of the enzymes should be heterodimers (AB). These would form a third, intermediate band between the A form (AA dimer) and the B form (BB dimer; fig. 5.9). The lack of heterodimers in the blood of heterozygotes is further proof that both G-6-PD alleles are not active within the same cells. That is, in any one cell, only AA or BB dimers can form, because no single cell has both the A and B forms.

The Lyon hypothesis has been demonstrated with many X-linked loci, but the most striking examples are those for color phenotypes in some mammals. For example, the tortoiseshell pattern of cats is due to the inactivation of X chromosomes (fig. 5.10). Tortoiseshell cats are normally females heterozygous for the *yellow* and *black* alleles of the X-linked color locus. They exhibit patches of these two colors, indicating that at a certain stage in development, one or the other of the X chromosomes was inactivated and all of the ensuing daughter cells in that line kept the same X chromosome inactive. The result is patches of coat color.

The X chromosome is inactivated starting at a point called the **X inactivation center (XIC).** That region contains a gene called *XIST* (for *X inactive-specific transcripts*, referring to the transcriptional activity of this gene in the inactivated X chromosome). The *XIST* gene has been putatively identified as the gene that initiates the inactivation of the X chromosome. This gene is known to be active only in the inactive X chromosome in a normal XX female. Another aspect of "Lyonization" is that several other loci are known to be active on the inactivated X chromosome; they are active in both X chromosomes, even though one is heterochromatic (inactivated). Although several of these loci are in the pseudoautosomal region of the short arm of the X chromosome, several other of the thirty or more genes known to be active are on other places on the mammalian X chromosome. Active genes on the inactive X include the gene for the enzyme steroid sulphatase; the red-cell antigen $Xg^a$; *MIC2;* a *ZFY*-like gene termed *ZFX;* the gene for Kallmann syndrome; and several others.

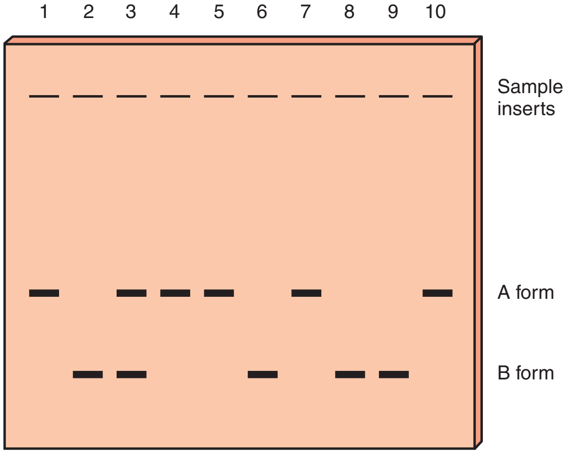


**Figure 5.8** Electrophoretic gel stained for glucose-6-phosphate dehydrogenase. Lanes *1–3* contain blood from an *AA* homozygote, a *BB* homozygote, and an *AB* heterozygote, respectively. Lanes *4–10* contain homogenates of individual cells of an *AB* heterozygote.

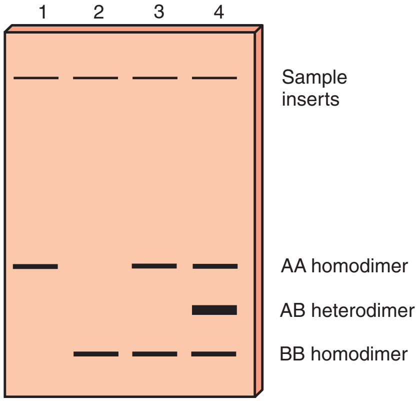


**Figure 5.9** Electrophoretic gel stained for glucose-6-phosphate dehydrogenase. Lanes *1* and *2* contain blood serum from *AA* and *BB* women, respectively, and lane *3* contains serum from an *AB* heterozygote. Lane *4* shows the pattern expected if both the *A* and *B* alleles were active within the same cell.

## BOX 5.2

### Experimental Methods

#### *Electrophoresis*

Electrophoresis, a technique for separating relatively similar types of molecules (for example, proteins and nucleic acids), has opened up new and exciting areas of research in population, biochemical, and molecular genetics. It has allowed us to see variations in large numbers of loci, previously difficult or impossible to sample. In biochemical genetics, electrophoretic techniques can be used to study enzyme pathways. In molecular genetics, electrophoresis is used to sequence nucleotides (see chapter 13) and to assign various loci to particular chromosomes. In population genetics (see chapter 21), electrophoresis has made it possible to estimate the amount of variability that occurs in natural populations.

Here we discuss protein electrophoresis, a process that entails placing a sample—often blood serum or a cell homogenate—at the top of a gel prepared from a suitable substrate (e.g., hydrolyzed starch, polyacrylamide, or cellulose acetate) and a buffer. An electrical current is

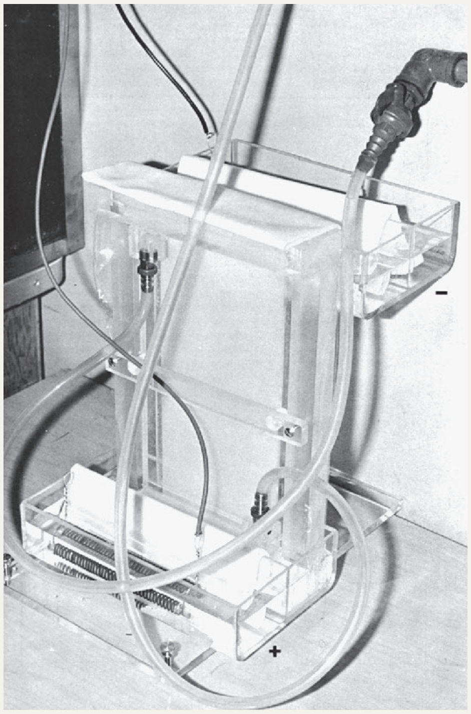


**Figure 1** Vertical starch gel apparatus. Current flows from the upper buffer chamber to the lower one by way of the paper wicks and the starch gel. Cooling water flows around the system. (R. P. Canham, "Serum protein variations and selection in fluctuating populations of cricetid rodents," Ph.D. thesis, University of Alberta, 1969. Reproduced by permission.)

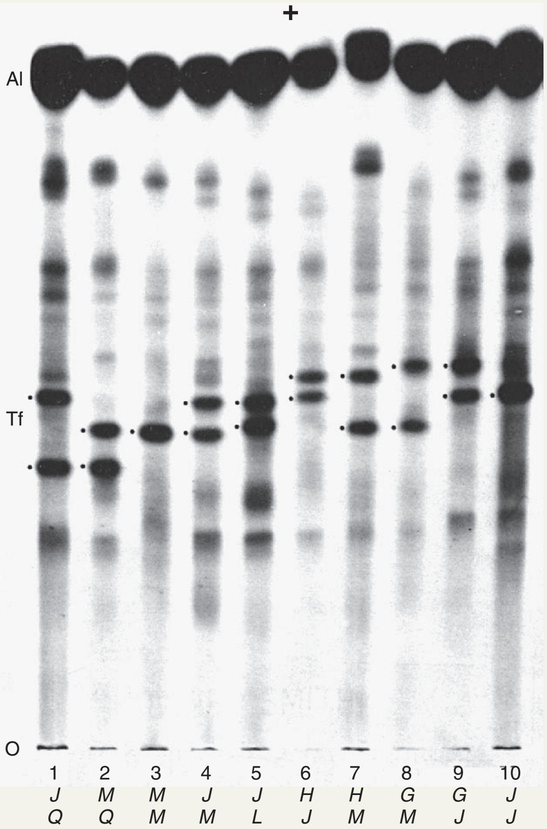


**Figure 2** Ten samples of deer mouse *(Peromyscus maniculatus)* blood studied for general protein. *Al* is albumin and *Tf* is transferrin, the two most abundant proteins in mammalian blood. The six *Tf* allozymes are labeled G, H, J, L, M, and Q. (R. P. Canham, "Serum protein variations and selection in fluctuating populations of cricetid rodents," Ph.D. thesis, University of Alberta, 1969. Reproduced by permission.)

passed through the gel to cause charged molecules to move (fig. 1), and the gel is then treated with a dye that stains the protein. In the simplest case, if a protein is homogeneous (usually the product of a homozygote), it forms a single band on the gel. If it is heterogeneous (usually the product of a heterozygote), it forms two bands. This is because the two allelic protein products differ by an amino acid; they have different electrical charges and therefore travel through the gel at different rates (see fig. 5.8). The term **allozyme** refers to different electrophoretic forms of an enzyme controlled by alleles at the same locus.

Figure 2 shows samples of mouse blood serum that have been stained for protein. Most of the staining reveals albumins and β-globulins (transferrin). Because they are present in very small concentrations, many enzymes present in the serum are not visible, but a stain that is specific for a particular enzyme can make that enzyme visible on the gel. For example, lactate dehydrogenase (LDH) can be located because it catalyzes this reaction:

$$\text{lactic acid} + \text{NAD}^+ \xleftrightarrow{\text{LDH}} \text{pyruvic acid} + \text{NADH}$$

Thus, we can stain specifically for the lactate dehydrogenase enzyme by adding the substrates of the enzyme (lactic acid and nicotinamide adenine dinucleotide, $\text{NAD}^+$) and a suitable stain specific for a product of the enzyme reaction (pyruvic acid or nicotinamide adenine dinucleotide, reduced form, NADH). That is, if lactic acid and $\text{NAD}^+$ are poured on the gel, only lactate dehydrogenase converts them to pyruvic acid and NADH. We can then test for the presence of NADH by having it reduce the dye, nitro blue tetrazolium, to the blue precipitate, formazan, an electron carrier. We then add all the preceding reagents and look for blue bands on the gel (fig. 3).

In addition to its uses in population genetics and chromosome mapping, electrophoresis has been extremely useful in determining the structure of many proteins and for studying developmental pathways. As we can see from the lactate dehydrogenase gel in figure 3, five bands can occur. In some tissues of a homozygote, these bands occur roughly in a ratio of 1:4:6:4:1. This pattern can come about if the enzyme is a tetramer whose four subunits are random mixtures of two gene products (from the *A* and *B* loci). Thus we would get

AAAA (1/16)
AAAB (4/16)
AABB (6/16)
ABBB (4/16)
BBBB (1/16)

(Note that the ratio 1:4:6:4:1 is the expansion of $[A + B]^4$, and the relative "intensity" of each band—the number of protein doses—is calculated from the rule of unordered events described in chapter 4.)

*continued*

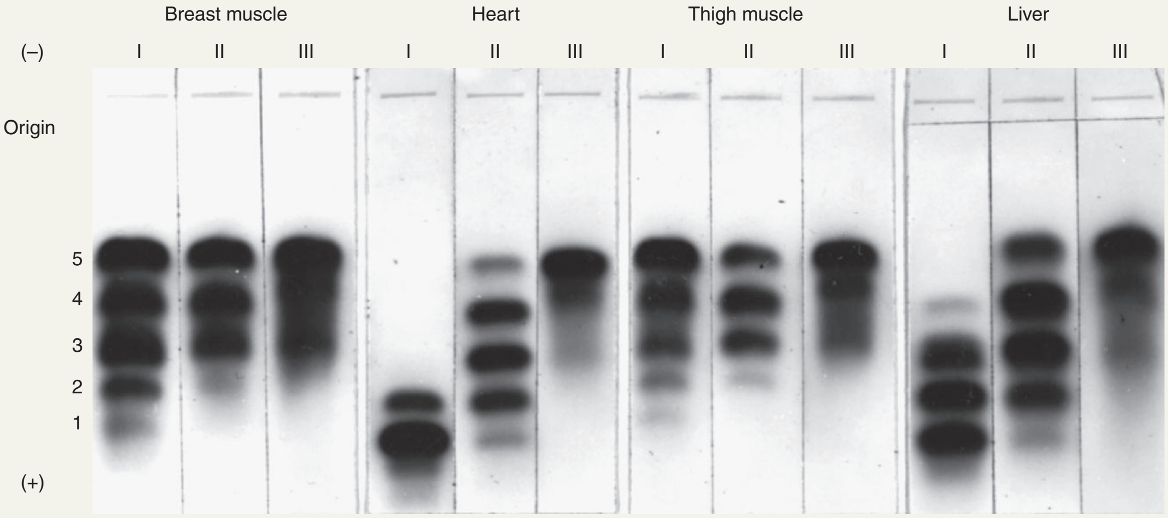


**Figure 3** Lactate dehydrogenase isozyme patterns in pigeons. Note the five bands for some individual samples. Lanes *I, II,* and *III* under each tissue type indicate the range of individual variation. (W. H. Zinkham, et al., "A Variant of Lactate Dehydrogenase in Somatic Tissues of Pigeons" in *Journal of Experimental Zoology* 162, no. 1 (June):45–46, 1966. Reproduced by permission of the Wistar Institute.)

## BOX 5.2 (CONTINUED)

Protein chemists have verified this tetramer model. In this way, electrophoresis has helped us determine the structure of several enzymes. (The term **isozymes** refers to multiple electrophoretic forms of an enzyme due to subunit interaction rather than allelic differences.)

Chemists have also discovered that the five forms differ in concentration in different tissues of the body (fig. 4). This has led to various hypotheses as to how the production of enzymes is controlled developmentally.

Electrophoresis is also valuable in clinical diagnosis. In various diseases, cell destruction causes the release of proteins into the bloodstream. Thus, the lactate dehydrogenase pattern is found in the blood in certain disease states (fig. 5). This is why examination of the blood LDH is often a diagnostic test used to pick up early signs of heart and liver disease (among others).

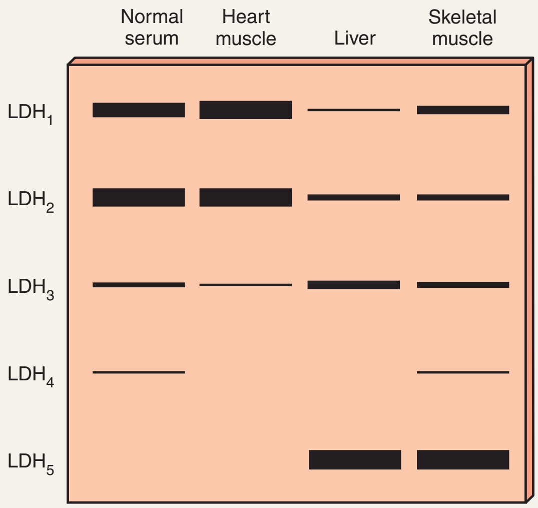


**Figure 4** LDH patterns found in different tissues in human beings.

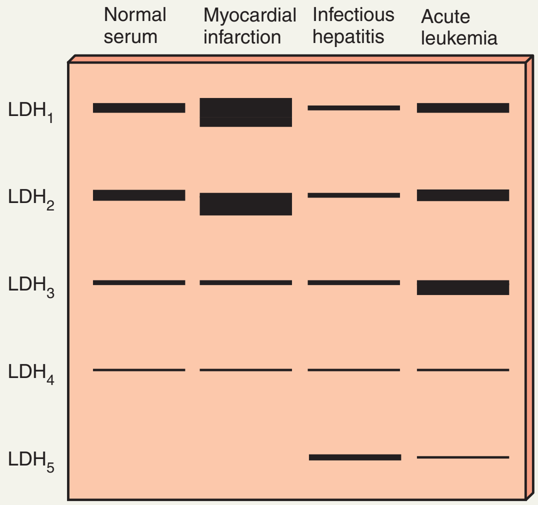


**Figure 5** LDH patterns from normal human serum and from serum affected by various disease states.

The gene product of *XIST* is an RNA that does not seem to be translated into a protein. Rather, using localization techniques, geneticists have found this RNA is associated with Barr bodies, coating the inactive chromosome. Current research is aimed at determining the details of this interaction.

## *Dosage Compensation for* Drosophila

Dosage compensation also occurs in fruit flies, and it appears that the gene activity of X chromosome loci is also about equal in males and females. The mechanism is different from that in mammals since no Barr bodies are found in fruit flies. Instead, the male's single X chromosome is hyperactive, approaching the level of transcriptional activity of both of the female's X chromosomes combined. Researchers have discovered a multisubunit protein complex called MSL (for male-specific lethal) that binds to hundreds of sites on the single X chromosome in males. Presumably, the binding mediates the hyperactivity of the genes on the X chromosome. (We discuss control of transcription later in the book.) At least five genes contribute products to this protein complex: *msl1, msl2, msl3, mle,* and *mof.* (*Mle* comes from *male*less, and *mof* comes from *m*ales absent *o*n the *f*irst.) Along with this protein complex are RNAs that also bind to the male X chromosome. These RNAs, also implicated in dosage compensation, are the products of the *rox1* and *rox2* genes (for *R*NA *o*n the *X*). Together, the MSL protein complex and the RNAs comprise a **compensasome.**

Mutant alleles of the male-specific lethal (*msl*) genes disrupt dosage compensation in males and are, as their names imply, lethal. However, they appear to have no effect in females. Expression of at least one of these genes, *msl2,* is repressed by the protein product of the *Sxl* gene. Thus, sex determination and dosage compensation are

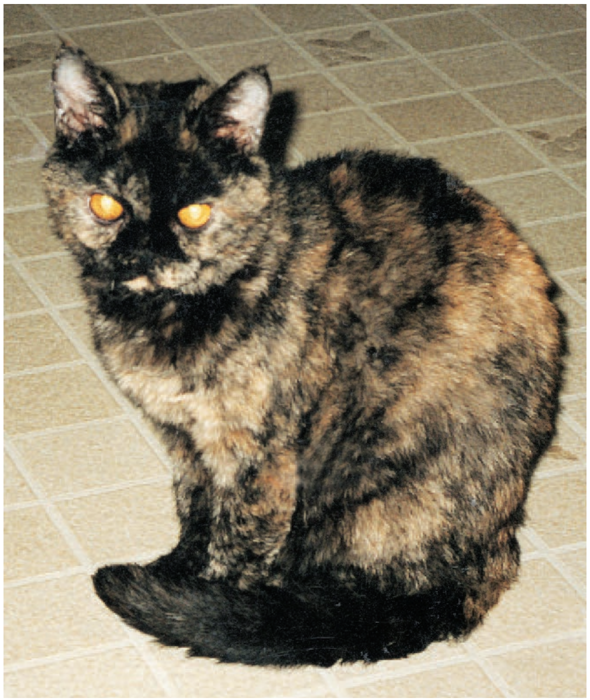


**Figure 5.10** Tortoiseshell cat. A female heterozygous for the X-linked *yellow* and *black* alleles. (Courtesy of Donna Bass.)

ultimately under the control of the same master switch gene, *Sxl.* This should not be surprising since the ability of *Sxl* to count the number of X chromosomes in a cell makes it the most efficient initiator of both sexual development and dosage compensation.

## SEX LINKAGE

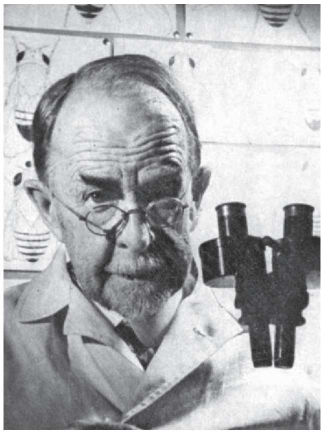
In an XY chromosomal system of sex determination, the pattern of inheritance for loci on the heteromorphic sex chromosomes differs from the pattern for loci on the homomorphic autosomal chromosomes because alleles of the sex chromosome are inherited in association with the sex of the offspring. Alleles on a male's X chromosome go to his daughters but not to his sons, because the presence of his X chromosome normally determines that his offspring is a daughter. For example, the inheritance pattern of hemophilia (failure of blood to clot), the common form of which is caused by an allele located on the X chromosome, has been known since the end of the eighteenth century. It was known that mostly men had the disease, whereas women could pass on the disease without actually having it. (In fact, the general nature of the inheritance of this trait was known in biblical times. The Talmud—the Jewish book of laws and traditions—specified exemptions to circumcision on the basis of hemophilia among relatives consistent with an understanding of who was at risk.)

Before we continue, we need to make a small distinction. Since both X and Y are sex chromosomes, three different patterns of inheritance are possible, all sex linked (for loci found only on the X chromosome, only on the Y chromosome, or on both). However, the term **sex-linked** usually refers to loci found only on the X chromosome; the term **Y-linked** is used to refer to loci found only on the Y chromosome, which control **holandric traits** (traits found only in males). Loci found on both the X and Y chromosomes are called *pseudoautosomal.* In human beings, at least four hundred loci are known to be on the X chromosome; only a few are known to be on the Y chromosome.

### *X Linkage in* Drosophila

T. H. Morgan demonstrated the **X-linked** pattern of inheritance in *Drosophila* in 1910, when a white-eyed male appeared in a culture of wild-type (red-eyed) flies (fig. 5.11). This male was crossed with a wild-type female. All of the offspring were wild-type. However, when these $F_1$ individuals were crossed with each other, their offspring fell into two categories (fig. 5.12). All the females and half the males were wild-type, whereas the remaining half of the males were white-eyed. Ultimately, Morgan and others interpreted this to mean that the white-eye locus was on the X chromosome. We can redraw figure 5.12 to include the sex chromosomes of Morgan's flies (fig. 5.13). We denote the X chromosome with the white-eye allele as $X^w$. Similarly $X^+$ is the X chromosome with the wild-type allele, and Y is the Y chromosome, which does not have this locus.

Another property of sex linkage appears in figure 5.13. Since females have two X chromosomes, they can have normal homozygous and heterozygous allelic combinations. But males, with only one copy of the X chromosome, can be neither homozygous nor heterozygous.



Thomas Hunt Morgan (1866–1945). (From *Genetics* 32 (1947): frontispiece. Courtesy of the Genetics Society of America.)

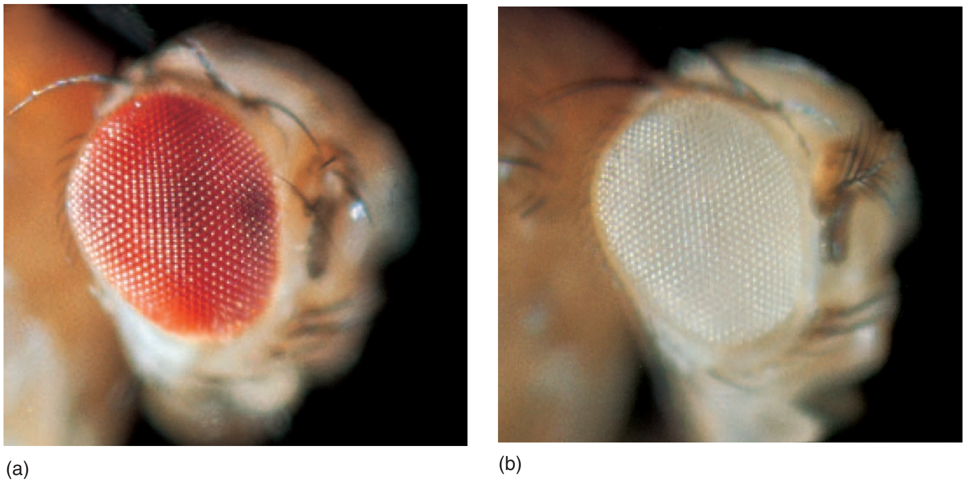

(a) (b)

**Figure 5.11** (*a*) Wild-type (red-eyed) and (*b*) white-eyed fruit flies. (Carolina Biological Supply Company.)

Instead, the term **hemizygous** describes the presence of X-linked genes (and other genes present in only one copy) in males. Since only one allele is present, a single copy of a recessive allele determines the phenotype in a phenomenon called **pseudodominance.** Thus, a male with one *w* allele is white-eyed, the allele acting in a dominant fashion. This is the same way one copy of a dominant autosomal allele would determine the phenotype of a normal diploid organism. Hence the term *pseudodominance.*

## Nonreciprocity

The X-linked pattern has long been known as the **crisscross pattern of inheritance** because the father passes a trait to his daughters, who pass it to their sons. Figure 5.14 shows why this analysis is correct and the inheritance pattern is not reciprocal through a cross between a white-eyed female and a wild-type male. Here the $F_1$ males are white-eyed, the $F_1$ females are wild-type, and 50% of each sex in the $F_2$ generation are white-eyed. Such nonreciprocity and different ratios in the two sexes suggest sex linkage, which the crisscross pattern confirms.

Figure 5.15 shows the inheritance pattern of a sex-linked trait in chickens, in which the male is the homogametic sex (ZZ). The gene for barred plumage is Z linked, and barred plumage is dominant to nonbarred plumage. If we substitute white-eyed for nonbarred and male for female, we get the same pattern as in fruit flies (fig. 5.13)—in which, of course, females are homogametic.

The Y chromosome in fruit flies carries the pseudoautosomal *bobbed* locus (*bb*), the nucleolar organizer. In the homozygous recessive state, it causes bristles to shorten. Figures 5.16 and 5.17 show the results of reciprocal crosses involving *bobbed.* In both cases, one quarter of the $F_2$ individuals are *bobbed.* In one cross it is males, and in the other it is females.

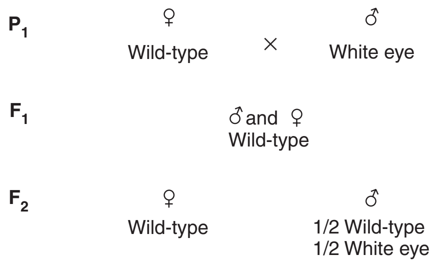


**Figure 5.12** Pattern of inheritance of the white-eye trait in *Drosophila.*

## *Sex-Limited and Sex-Influenced Traits*

Aside from X-linked, holandric, and pseudoautosomal inheritance, two inheritance patterns show nonreciprocity without necessarily being under the control of loci on the sex chromosomes. **Sex-limited traits** are traits expressed in only one sex, although the genes are present in both. In women, breast and ovary formation are sex-limited traits, as are facial hair distribution and sperm production in men. Nonhuman examples are plumage patterns in birds—in many species, the male is brightly colored—and horns found only in males of certain sheep species. Milk yield in mammals is expressed phenotypically only in females. **Sex-influenced,** or **sex-conditioned, traits** appear in both sexes but occur in one sex more than the other. Pattern, or premature, baldness in human beings is an example of a sex-influenced trait. In women, it is usually expressed as a thinning of hair rather than as balding. Apparently testosterone, the male hormone, is required for the full expression of the allele.

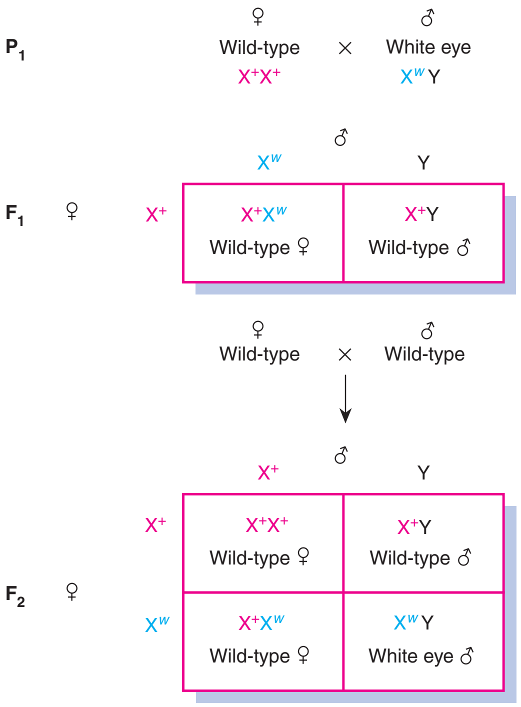


**Figure 5.13** Crosses of figure 5.12 redrawn to include the sex chromosomes.

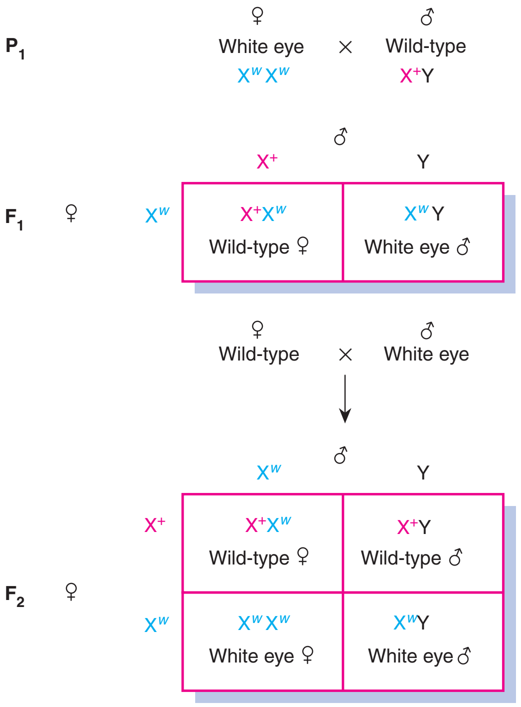


**Figure 5.14** Reciprocal cross to that in figure 5.13.

## PEDIGREE ANALYSIS

Inheritance patterns in many organisms are relatively easy to determine, because crucial crosses can test hypotheses about the genetic control of a particular trait. Many of these same organisms produce an abundance of offspring so that investigators can gather numbers large enough to compute ratios. Recall Mendel's work with garden peas; his 3:1 ratio in the $F_2$ generation led him to suggest the rule of segregation. If Mendel's sample sizes had been smaller, he might not have seen the ratio. Think of the difficulties Mendel would have faced had he decided to work with human beings instead of pea plants. Human geneticists face the same problems today. The occurrence of a trait in one of four children does not necessarily indicate a true 3:1 ratio.

To determine the inheritance pattern of many human traits, human geneticists often have little more to go on than a **pedigree** that many times does not include critical mating combinations. Frequently uncertainties and ambiguities plague pedigree analysis, a procedure whereby conclusions are often a product of the process of elimination. Other difficulties human geneticists encounter are the lack of **penetrance** and different degrees of **expressivity** in many traits. Both are aspects of the expression of a phenotype.

### *Penetrance and Expressivity*

*Penetrance* refers to the appearance in the phenotype of genotypically determined traits. Unfortunately for geneticists, not all genotypes "penetrate" the phenotype. For example, a person could have the genotype that specifies vitamin-D-resistant rickets and yet not have rickets (a bone disease). This disease is caused by a sex-linked dominant allele and is distinguished from normal vitamin D deficiency by its failure to respond to low levels of vitamin D. It does, however, respond to very high levels of vitamin D and is thus treatable. In any case, in some family trees, affected children are born to unaffected parents. This would violate the rules of dominant inheritance because one of the parents must have had the allele yet did not express it. The fact that the parent actually had the allele is demonstrated by the occurrence of low levels of phosphorus in the blood, a pleiotropic effect of the same allele. The low-phosphorus aspect of the phenotype is always fully penetrant.

Thus, certain genotypes, often those for developmental traits, are not always fully penetrant. Most genotypes, however, are fully penetrant. For example, no known cases exist of individuals homozygous for albinism who do not actually lack pigment. Vitamin-D-resistant rickets illustrates another case in which a phenotype that is not genetically determined mimics a phenotype that is. This

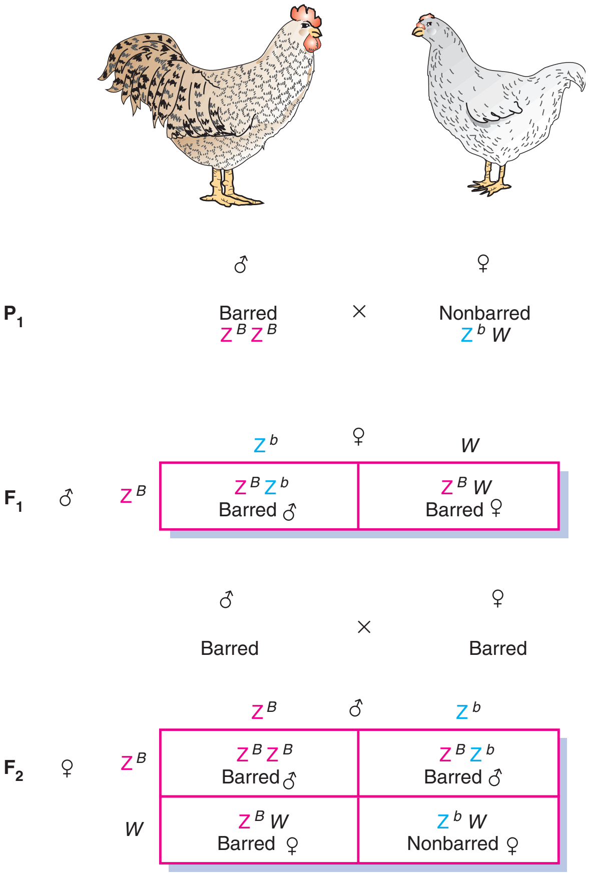


**Figure 5.15** Inheritance pattern of barred plumage in chickens in which males are homogametic (ZZ) and females are heterogametic (ZW).

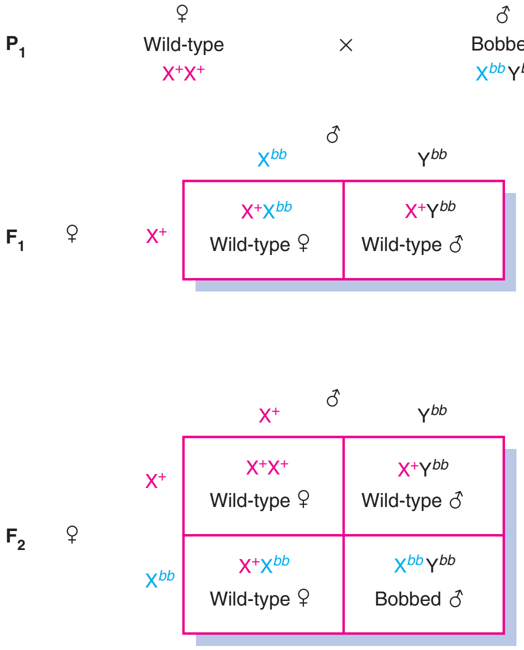


**Figure 5.16** Inheritance pattern of the *bobbed* locus in *Drosophila.*

**phenocopy** is the result of dietary deficiency or environmental trauma. A dietary deficiency of vitamin D, for example, produces rickets that is virtually indistinguishable from genetically caused rickets.

Many developmental traits not only sometimes fail to penetrate, but also show a variable pattern of expression, from very mild to very extreme, when they do. For example, cleft palate is a trait that shows both variable penetrance and variable expressivity. Once the genotype penetrates, the severity of the impairment varies considerably, from a very mild external cleft to a very severe clefting of the hard and soft palates. Failure to penetrate and variable expressivity are not unique to human traits but are characteristic of developmental traits in many organisms.

## *Family Tree*

One way to examine a pattern of inheritance is to draw a family tree. Figure 5.18 defines the symbols used in constructing a family tree, or pedigree. The circles represent females, and the squares represent males. Symbols that are filled in represent individuals who have the trait under study; these individuals are said to be **affected.** The open symbols represent those who do not have the trait. Direct horizontal lines between two individuals (one male, one female) are called marriage lines. Children are attached to a marriage line by a vertical line. All the brothers and sisters **(siblings** or **sibs)** from the same parents are connected by a horizontal line above their symbols. Siblings are numbered below their symbols according to birth order (fig. 5.19), and generations are numbered on the right in Roman numerals. When the sex of a child is unknown, the symbol is diamond-shaped (e.g., the children of III-1 and III-2 in fig. 5.19). A number within a symbol represents the number of siblings not separately listed. Individuals IV-7 and IV-8 in figure 5.19 are fraternal (dizygotic or nonidentical) twins: they originate from the same point. Individuals III-3 and III-4 are identical (monozygotic) twins: they originate from the same short vertical line.

When other symbols occur in a pedigree, they are usually defined in the legend. Individual V-5 in figure 5.19 is called a **proband** or **propositus** (female, **proposita**). The arrow pointing to individual V-5 indicates that the pedigree was ascertained through this individual, usually by a physician or clinical investigator.

On the basis of the information in a pedigree, geneticists attempt to determine the mode of inheritance

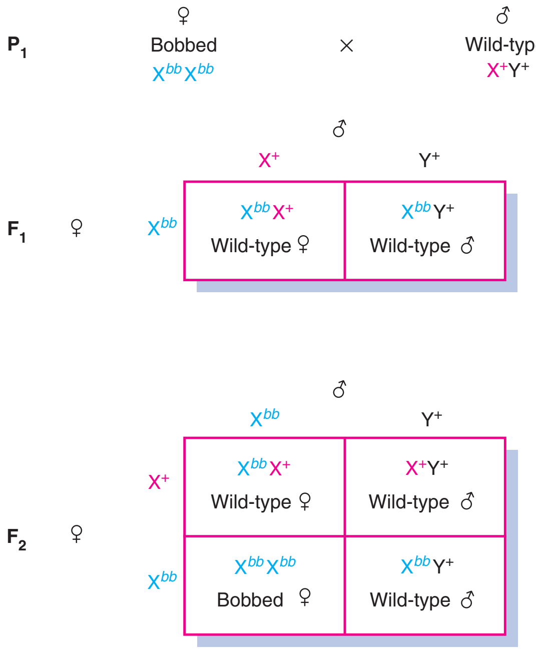


**Figure 5.17** Reciprocal cross to that in figure 5.16.

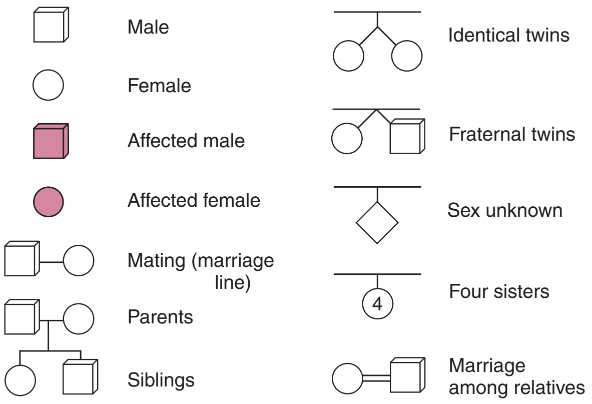


**Figure 5.18** Symbols used in a pedigree.

of a trait. There are two types of questions the pedigree might be used to answer. First, are there patterns within the pedigree that are consistent with a particular mode of inheritance? Second, are there patterns within the pedigree that are inconsistent with a particular mode of inheritance? Often, it is not possible to determine the mode of inheritance of a particular trait with certainty. McKusick has reported that, as of 2001, the mode of inheritance of over nine thousand loci in human beings was known with some confidence, including autosomal dominant, autosomal recessive, and sex-linked genes.

## Dominant Inheritance

If we look again at the pedigree in figure 5.19, several points emerge. First, polydactyly (fig. 5.20) occurs in every generation. Every affected child has an affected parent—no generations are skipped. This suggests dominant inheritance. Second, the trait occurs about equally in both sexes; there are seven affected males and six affected females in the pedigree. This indicates autosomal rather than sex-linked inheritance. Thus, so far, we would categorize polydactyly as an autosomal dominant trait. Note also that individual IV-11, a male, passed on the trait to two of his three sons. This would rule out sex linkage. (Remember that a male gives his X chromosome to all of his daughters but none of his sons. His sons receive his Y chromosome.) Consistency in many such pedigrees, has confirmed that an autosomal dominant gene causes polydactyly.

Polydactyly shows variable penetrance and expressivity. The most extreme manifestation of the trait is an extra digit on each hand (fig. 5.20) and one or two extra toes on each foot. However, some individuals have only extra toes, some have extra fingers, and some have an asymmetrical distribution of digits such as six toes on one foot and seven on the other.

## Recessive Inheritance

Figure 5.21 is a pedigree with a different pattern of inheritance. Here affected individuals are not found in each generation. The affected daughters, identical triplets, come from unaffected parents. They represent, in fact, the first appearance of the trait in the pedigree. A telling point is that the parents of the triplets are first cousins; a mating between relatives is referred to as **consanguineous.** If the degree of relatedness is closer than law permits, the union is called **incestuous.** In all states, brother-sister and mother-son marriages are forbidden; and in all states except Georgia, father-daughter marriages are forbidden. Georgia did not intend to permit father-daughter marriages. However, the law was drafted using biblical terminology that inadvertently did not prohibit a man from marrying his daughter or his grandmother. Thirty states prohibit the marriage of first cousins.

Consanguineous matings often produce offspring that have rare recessive, and often deleterious, traits. The reason is that through common ancestry (e.g., when first cousins have a pair of grandparents in common), a rare allele can be passed on both sides of the pedigree and become homozygous in a child. The occurrence of a trait in a pedigree with common ancestry is often good evidence

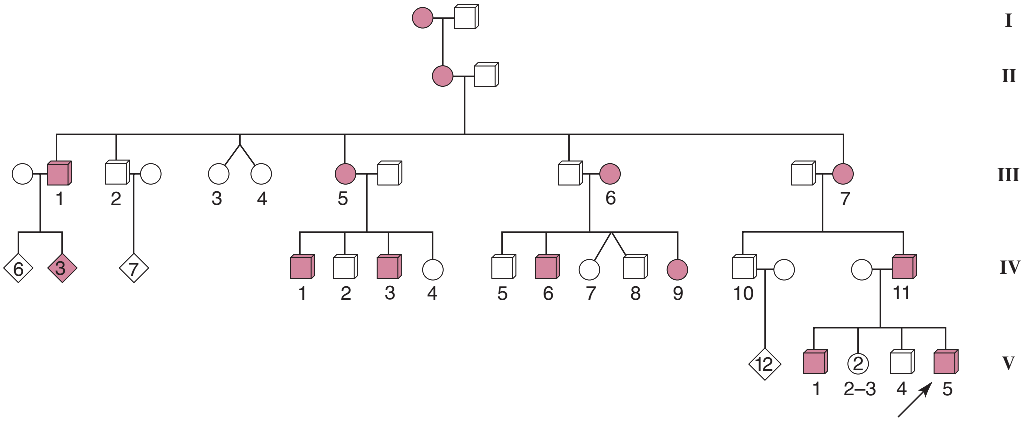


**Figure 5.19** Part of a pedigree for polydactyly.

for an autosomal recessive mode of inheritance. Consanguinity by itself does not guarantee that a trait has an autosomal recessive mode of inheritance; all modes of inheritance appear in consanguineous pedigrees. Conversely, recessive inheritance is not confined to consanguineous pedigrees. Hundreds of recessive traits are known from pedigrees lacking consanguinity.

## *Sex-Linked Inheritance*

Figure 5.22 is the pedigree of Queen Victoria of England. Through her children, hemophilia was passed on to many of the royal houses of Europe. Several interesting aspects of this pedigree help to confirm the method of inheritance. First, generations are skipped. Although Alexis (1904–18) was a hemophiliac, neither his parents nor his grandparents were. This pattern occurs in several other places in the pedigree and indicates a recessive mode of inheritance. From other pedigrees and from the biochemical nature of the defect, scientists have determined that hemophilia is a recessive trait.

Further inspection of the pedigree in figure 5.22 reveals that all the affected individuals are sons, strongly suggesting sex linkage. Since males are hemizygous for the X chromosome, more males than females should have the phenotype of a sex-linked recessive trait because males do not have a second X chromosome that might carry the normal allele. If this is correct, we can make several predictions. First, since all males get their X chromosomes from their mothers, affected males should be the offspring of carrier (heterozygous) females. A female is automatically a carrier if her father had the disease. She has a 50% chance of being a carrier if her brother, but not her father, has the disease. In that case, her mother was a carrier. The pedigree in figure 5.22 is consistent with these predictions.

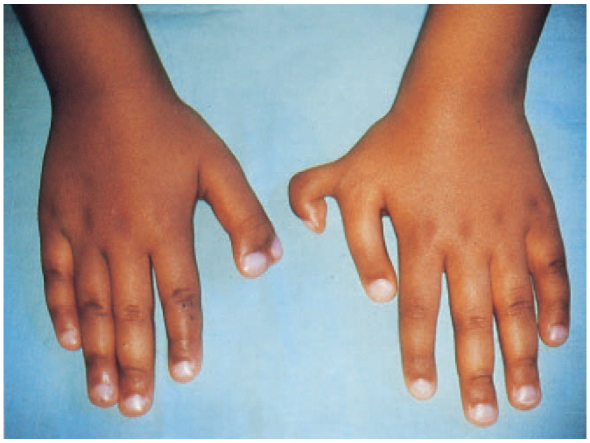


**Figure 5.20** Hands of a person with polydactyly. Manifestations range in severity from one extra finger or toe to one or more extra digits on each hand and foot. (© L.V. Bergman/The Bergman Collection.)

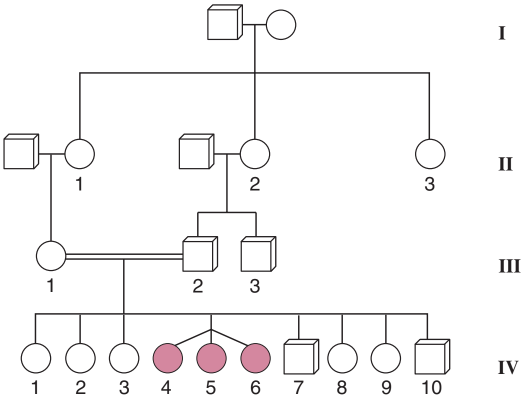
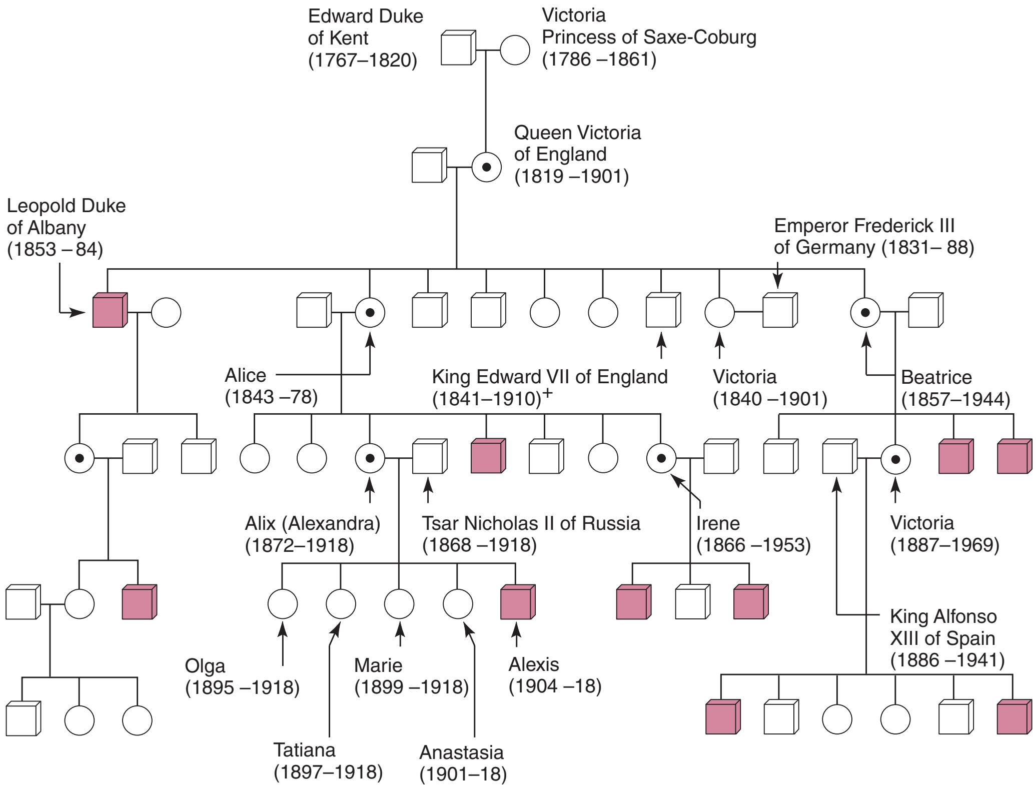


**Figure 5.21** Part of a pedigree of hypotrichosis (early hair loss).


Edward Duke of Kent (1767–1820)
Victoria Princess of Saxe-Coburg (1786 –1861)
Queen Victoria of England (1819 –1901)
Leopold Duke of Albany (1853 – 84)
Emperor Frederick III of Germany (1831– 88)
Alice (1843 –78)
King Edward VII of England (1841–1910)+
Victoria (1840 –1901)
Beatrice (1857–1944)
Alix (Alexandra) (1872–1918)
Tsar Nicholas II of Russia (1868 –1918)
Irene (1866 –1953)
Victoria (1887–1969)
King Alfonso XIII of Spain (1886 –1941)
Olga (1895 –1918)
Marie (1899 –1918)
Alexis (1904 –18)
Tatiana (1897–1918)
Anastasia (1901–18)

○ Normal female

⊙ Normal, but known carrier (heterozygous) female

☐ Normal male

🟥 Affected male

+ Descendants include present British royal family

**Figure 5.22** Hemophilia in the pedigree of Queen Victoria of England. In the photograph of the Queen and some of her descendants, three carriers—Queen Victoria (*center*), Princess Irene of Prussia (*right*), and Princess Alix (Alexandra) of Hesse (*left*)—are indicated. (Photo © Mary Evans Picture Library/Photo Researchers, Inc.)

In no place in the pedigree is the trait passed from father to son. This would defy the route of an affected X chromosome. We can conclude from the pedigree that hemophilia is a sex-linked recessive trait. (Several different inherited forms of hemophilia are known, each deficient in one of the steps in the pathway that forms fibrinogen, the blood clot protein. Two of these forms, "classic" hemophilia A and hemophilia B, also called Christmas disease, are sex linked. Other hemophilias are autosomal.)

One other interesting point about this pedigree is that there is no evidence of the disease in Queen Victoria's ancestors, yet she was obviously a heterozygote, having one affected son and two daughters who were known carriers. Thus, though she was born to what appears to be a homozygous normal mother and a hemizygous normal father, one of Queen Victoria's X chromosomes had the hemophilia allele. This could have happened if a change (mutation) had occurred in one of the gametes that formed Queen Victoria. (We explore the mechanisms of mutation in chapter 12.)

Figure 5.23 is another pedigree that points to dominant inheritance because the trait skips no generations. The pedigree shows the distribution of low blood-phosphorus levels, the fully penetrant aspect of vitamin-D-resistant rickets, among the sexes. Affected males pass on the trait to their daughters but not their sons. This pattern follows that of the X chromosome: a male passes it on to all of his daughters but to none of his sons. Although this pedigree accords with a sex-linked dominant mode of inheritance, it does not rule out autosomal inheritance. The pedigree shown is a small part of one involving hundreds of people, all with phenotypes consistent with the hypothesis of sex-linked dominant inheritance.

In figure 5.23, there is the slight possibility that the trait is recessive. This could be true if the male in generation I and the mates of II-5 and II-7 were all heterozygotes. Since this is a rare trait, the possibility that all these conditions occurred is small. For example, if one person in fifty (0.02) is a heterozygote, then the probability of three heterozygotes mating within the same pedigree is $(0.02)^3$, or eight in one million. The rareness of this event further supports the hypothesis of dominant inheritance. The expected patterns for the various types of inheritance in pedigrees can be summarized in the following four categories:

### Autosomal Recessive Inheritance

1. Trait often skips generations.
2. An almost equal number of affected males and females occur.
3. Traits are often found in pedigrees with consanguineous matings.
4. If both parents are affected, all children should be affected.
5. In most cases when unaffected people mate with affected individuals, all children are unaffected. When at least one child is affected (indicating that the unaffected parent is heterozygous), approximately half the children should be affected.
6. Most affected individuals have unaffected parents.

### Autosomal Dominant Inheritance

1. Trait should not skip generations (unless trait lacks full penetrance).
2. When an affected person mates with an unaffected person, approximately 50% of their offspring should be affected (indicating also that the affected individual is heterozygous).
3. The trait should appear in almost equal numbers among the sexes.

### Sex-Linked Recessive Inheritance

1. Most affected individuals are male.
2. Affected males result from mothers who are affected or who are known to be carriers (heterozygotes) because they have affected brothers, fathers, or maternal uncles.
3. Affected females come from affected fathers and affected or carrier mothers.
4. The sons of affected females should be affected.
5. Approximately half the sons of carrier females should be affected.

### Sex-Linked Dominant Inheritance

1. The trait does not skip generations.
2. Affected males must come from affected mothers.
3. Approximately half the children of an affected heterozygous female are affected.
4. Affected females come from affected mothers or fathers.
5. All the daughters, but none of the sons, of an affected father are affected.

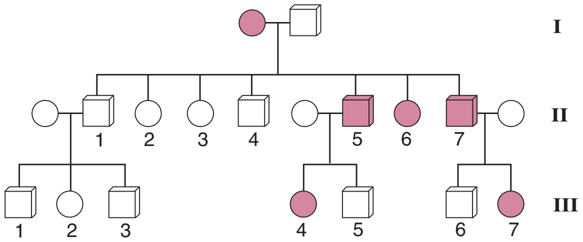


**Figure 5.23** Part of a pedigree of vitamin-D-resistant rickets. Affected individuals have low blood-phosphorus levels. Although the sample is too small for certainty, dominance is indicated because every generation was affected, and sex linkage is suggested by the distribution of affected individuals.

# SUMMARY

This chapter begins a four-chapter sequence that analyzes the relationship of genes to chromosomes. We begin with the study of sex determination.

**STUDY OBJECTIVE 1:** To analyze the causes of sex determination in various organisms 83–90

Sex determination in animals is often based on chromosomal differences. In human beings and fruit flies, females are homogametic (XX) and males are heterogametic (XY). In human beings, a locus on the Y chromosome, *SRY,* determines maleness; in *Drosophila,* sex is determined by the balance between genes on the X chromosome and genes on the autosomes that regulate the state of the sex-switch gene, *Sxl.*

**STUDY OBJECTIVE 2:** To understand methods of dosage compensation 90–95

Different organisms have different ways of solving problems of dosage compensation for loci on the X chromosome. In human beings, one of the X chromosomes in cells in a woman is Lyonized, or inactivated. Lyonization in women leads to cellular mosaicism for most loci on the X chromosome. In *Drosophila,* the X chromosome in males is hyperactive.

**STUDY OBJECTIVE 3:** To analyze the inheritance patterns of traits that loci on the sex chromosomes control 95–97

Since different chromosomes are normally associated with each sex, inheritance of loci located on these chromosomes shows specific, nonreciprocal patterns. The white-eye locus in *Drosophila* was the first case when a locus was assigned to the X chromosome. Over four hundred sex-linked loci are now known in human beings.

**STUDY OBJECTIVE 4:** To use pedigrees to infer inheritance patterns 97–102

Human genetic studies use pedigree analysis to determine inheritance patterns because it is impossible to carry out large-scale, controlled human crosses. However, not all traits determined by genotype are apparent in the phenotype, and this lack of penetrance can pose problems in genetic analysis.

# SOLVED PROBLEMS

**PROBLEM 1:** A Female fruit fly with a yellow body is discovered in a wild-type culture. The female is crossed with a wild-type male. In the $F_1$ generation, the males are yellow-bodied and the females are wild-type. When these flies are crossed among themselves, the $F_2$ produced are both yellow-bodied and wild-type, equally split among males and females (see fig. 1). Explain the genetic control of this trait.

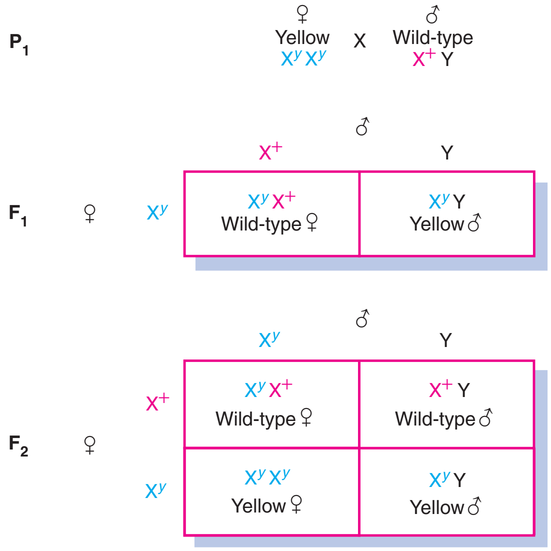


**Figure 1** Cross between yellow-bodied and wild-type fruit flies.

***Answer:*** Since the results in the $F_1$ generation differ between the two sexes, we suspect that a sex-linked locus is responsible for the control of body color. If we assume that it is a recessive trait, then the female parent must have been a recessive homozygote, and the male must have been a wild-type hemizygote. If we assign the wild-type allele as $X^+$, the yellow-body allele as $X^y$, and the Y chromosome as Y, then figure 1, showing the crosses into the $F_2$ generation, is consistent with the data. Thus, a recessive X-linked gene controls yellow body color in fruit flies.

**PROBLEM 2:** The affected individuals in the pedigree in figure 2 are chronic alcoholics (data from the National Institute of Alcohol Abuse and Alcoholism). What can you say about the inheritance of this trait?

***Answer:*** We begin by assuming 100% penetrance. If that is the case, then we can rule out either sex-linked or autosomal recessive inheritance because both parents had the trait, yet they produced some unaffected children.

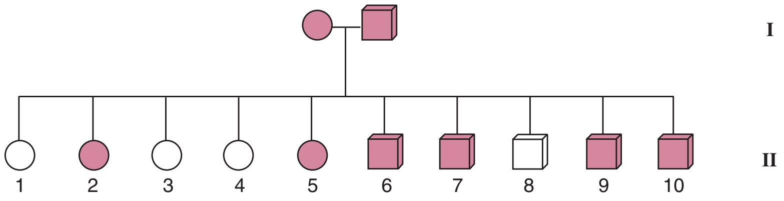


**Figure 2** A pedigree for alcoholism.

Nor can the mode of inheritance be by a sex-linked dominant gene because an affected male would have only affected daughters, since his daughters get copies of his single X chromosome. We are thus left with autosomal dominance as the mode of inheritance. If that is the case, then both parents must be heterozygotes; otherwise, all the children would be affected. If both parents are heterozygotes, we expect a 3:1 ratio of affected to unaffected offspring (a cross of $Aa \times Aa$ produces offspring of $A\text{-}:aa$ in a 3:1 ratio); here, the ratio is 6:4. If we did a chi-square test, the expected numbers would be 7.5:2.5 (3/4 and 1/4, respectively, of 10). Although the expected value of 2.5 makes it inappropriate to do a chi-square test (the expected value is too small), we can see that the observed and expected numbers are very close. Thus, from the pedigree we would conclude that an autosomal dominant allele controls chronic alcoholism. (Although the analysis is consistent, we actually cannot draw that conclusion about alcoholism because other pedigrees are not consistent with 100% penetrance, a one-gene model, or the lack of environmental influences. In fact, scientists are currently debating whether alcoholism is inherited at all. These types of problems related to complex human traits are discussed in chapter 18.)

**PROBLEM 3:** A female fly with orange eyes is crossed with a male fly with short wings. The $F_1$ females have wild-type (red) eyes and long wings; the $F_1$ males have orange eyes and long wings. The $F_1$ flies are crossed to yield

47 long wings, red eyes
45 long wings, orange eyes
17 short wings, red eyes
14 short wings, orange eyes

with no differences between the sexes. What is the genetic basis of each trait?

***Answer:*** In the $F_1$ flies, we see a difference in eye color between the sexes, indicating some type of sex linkage. Since the females are wild-type, wild-type is probably dominant to orange. We can thus diagram the cross for eye color as

$$\begin{array}{cccc} \text{(female) } X^oX^o & & X^+Y \text{ (male)} & P_1 \\ \text{orange} & & \text{red} & \\ & \downarrow & & \\ X^+X^o & & X^oY & F_1 \\ \text{red} & & \text{orange} & \\ & \downarrow & & \\ X^+X^o \quad X^oX^o & & X^+Y \quad X^oY & F_2 \\ \text{red} \quad \text{orange} & & \text{red} \quad \text{orange} & \end{array}$$

We would thus expect to see equal numbers of red-eyed and orange-eyed males and females, which is what we observe. Now look at long versus short wings. If we disregard eye color, wing length seems to be under autosomal control with short wings being recessive. Thus, the parents are homozygotes ($ss$ and $s^+s^+$), the $F_1$ offspring are heterozygotes ($s^+s$), and the $F_2$ progeny have a phenotypic ratio of 3:1, wild-type (long) to short wings.

## EXERCISES AND PROBLEMS*

### SEX DETERMINATION

1. What is the difference between an X and a Z chromosome?
2. Transformer (*tra*) is an autosomal recessive gene that converts chromosomal females into sterile males. A female *Drosophila* heterozygous for the transformer allele (*tra*) is mated with a normal male homozygous for transformer. What is the sex ratio of their offspring? What is the sex ratio of their offspring's offspring?
3. The autosomal recessive doublesex (*dsx*) gene converts males and females into developmental intersexes. Two fruit flies, both heterozygous for the *doublesex* (*dsx*) allele, are mated. What are the sexes of their offspring?
4. What is a sex switch? What genes serve as sex switches in human beings and *Drosophila?*

---

*Answers to selected exercises and problems are on page A-4.

## DOSAGE COMPENSATION

5. The diagram of an electrophoretic gel in figure 3 shows activity for a particular enzyme. Lane 1 is a sample from a "fast" homozygote. Lane 2 is a sample from a "slow" homozygote. In lane 3, the blood from the first two was mixed. Lane 4 comes from one of the children of the two homozygotes.

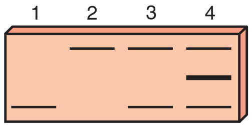


**Figure 3** The activity of a particular enzyme as revealed on an electrophoretic gel.

Can you guess the structure of the enzyme? If this were an X-linked trait, what pattern would you expect from a heterozygous female's
   - **a.** whole blood?
   - **b.** individual cells?

6. How many different zones of activity (bands) would you see on a gel stained for lactate dehydrogenase (LDH) activity from a person homozygous for the A protein gene but heterozygous for the B protein gene? Are the bands due to the activity of allozymes or isozymes?

7. How many Barr bodies would you see in the nuclei of persons with the following sex chromosomes?
   - **a.** X0
   - **b.** XX
   - **c.** XY
   - **d.** XXY
   - **e.** XXX
   - **f.** XXXXX
   - **g.** XX/XY mosaic

   What would the sex of each of these persons be? If these were the sex chromosomes of individual *Drosophila* that were diploid for all other chromosomes, what would their sexes be?

8. Calico cats have large patches of colored fur. What does this indicate about the age of onset of Lyonization (is it early or late)? Tortoiseshell cats have very small color patches. Explain the difference between the two phenotypes.

## SEX LINKAGE

9. In *Drosophila,* the lozenge phenotype, caused by a sex-linked recessive allele (*lz*), is narrow eyes. Diagram to the $F_2$ generation a cross of a lozenge male and a homozygous normal female. Diagram the reciprocal cross.

10. Sex linkage was originally detected in 1906 in moths with a ZW sex-determining mechanism. In the currant moth, a pale color (*p*) is recessive to the wild-type and located on the Z chromosome. Diagram reciprocal crosses to the $F_2$ generation in these moths.

11. What family history of hemophilia would indicate to you that a newborn male baby should be exempted from circumcision?

12. What is the difference between *pseudodominance* and *phenocopy?*

13. In *Drosophila,* cut wings are controlled by a recessive sex-linked allele (*ct*), and fuzzy body is controlled by a recessive autosomal allele (*fy*). When a fuzzy female is mated with a cut male, all the members of the $F_1$ generation are wild-type. What are the proportions of $F_2$ phenotypes, by sex?

14. Consider the following crosses in canaries:

| **Parents** | **Progeny** |
|---|---|
| **a.** pink-eyed female × pink-eyed male | all pink-eyed |
| **b.** pink-eyed female × black-eyed male | all black-eyed |
| **c.** black-eyed female × pink-eyed male | all females pink-eyed,<br>all males black-eyed |

Explain these results by determining which allele is dominant and how eye color is inherited.

15. Consider the following crosses involving yellow and gray true-breeding *Drosophila:*

| **Cross** | **$F_1$** | **$F_2$** |
|---|---|---|
| gray female × yellow male | all males gray,<br>all females gray | 97 gray females,<br>42 yellow males,<br>48 gray males |
| yellow female × gray male | all females gray,<br>all males yellow | ? |

   - **a.** Is color controlled by an autosomal or an X-linked gene?
   - **b.** Which allele, gray or yellow, is dominant?
   - **c.** Assume 100 $F_2$ offspring are produced in the second cross. What kinds and what numbers of progeny do you expect? List males and females separately.

16. A man with brown teeth mates with a woman with normal white teeth. They have four daughters, all with brown teeth, and three sons, all with white teeth. The sons all mate with women with white teeth, and all their children have white teeth. One of the daughters (A) mates with a man with white teeth (B), and they have two brown-toothed daughters, one white-toothed daughter, one brown-toothed son, and one white-toothed son.
   - **a.** Explain these observations.
   - **b.** Based on your answer to *a,* what is the chance that the next child of the A-B couple will have brown teeth?

17. In human beings, red-green color blindness is inherited as an X-linked recessive trait. A woman with normal vision whose father was color-blind marries a man with normal vision whose father was also color-

blind. This couple has a color-blind daughter with a normal complement of chromosomes. Is infidelity suspected? Explain.

**18.** A white-eyed male fly is mated with a pink-eyed female. All the $F_1$ offspring have wild-type red eyes. $F_1$ individuals are mated among themselves to yield:

| **Females** | | **Males** | |
|---|---|---|---|
| red-eyed | 450 | red-eyed | 231 |
| pink-eyed | 155 | white-eyed | 301 |
| | | pink-eyed | 70 |

Provide a genetic explanation for the results.

**19.** In *Drosophila,* white eye is an X-linked recessive trait, and ebony body is an autosomal recessive trait. A homozygous white-eyed female is crossed with a homozygous ebony male.

**a.** What phenotypic ratio do you expect in the $F_1$ generation?

**b.** What phenotypic ratio do you expect in the $F_2$ generation?

**c.** Suppose the initial cross was reversed: ebony female × white-eyed male. What phenotypic ratio would you expect in the $F_2$ generation?

**20.** In *Drosophila,* abnormal eyes can result from mutations in many different genes. A true-breeding wild-type male is mated with three different females, each with abnormal eyes. The results of these crosses are as follows:

| | | **Females** | **Males** |
|---|---|---|---|
| male × abnormal-1 | → | all normal | all normal |
| male × abnormal-2 | → | 1/2 normal,<br>1/2 abnormal | 1/2 normal,<br>1/2 abnormal |
| male × abnormal-3 | → | all abnormal | all abnormal |

Explain the results by determining the mode of inheritance for each abnormal trait.

**21.** A black and orange female cat is crossed with a black male, and the progeny are as follows:

females: two black, three orange and black

males: two black, two orange

Explain the results.

**22.** Based on the following *Drosophila* crosses, explain the genetic basis for each trait and determine the genotypes of all individuals:

white-eyed, dark-bodied female × red-eyed, tan-bodied male

$F_1$: females are all red-eyed, tan-bodied; males are all white-eyed, tan-bodied

$F_2$: 27 red-eyed, tan-bodied
24 white-eyed, tan-bodied
9 red-eyed, dark-bodied
7 white-eyed, dark-bodied

(No differences between males and females in the $F_2$ generation.)

## PEDIGREE ANALYSIS

**23.** What is the difference between penetrance and expressivity?

**24.** What are the possible modes of inheritance in pedigrees *a–c* in figure 4? What modes of inheritance are not possible for a given pedigree?

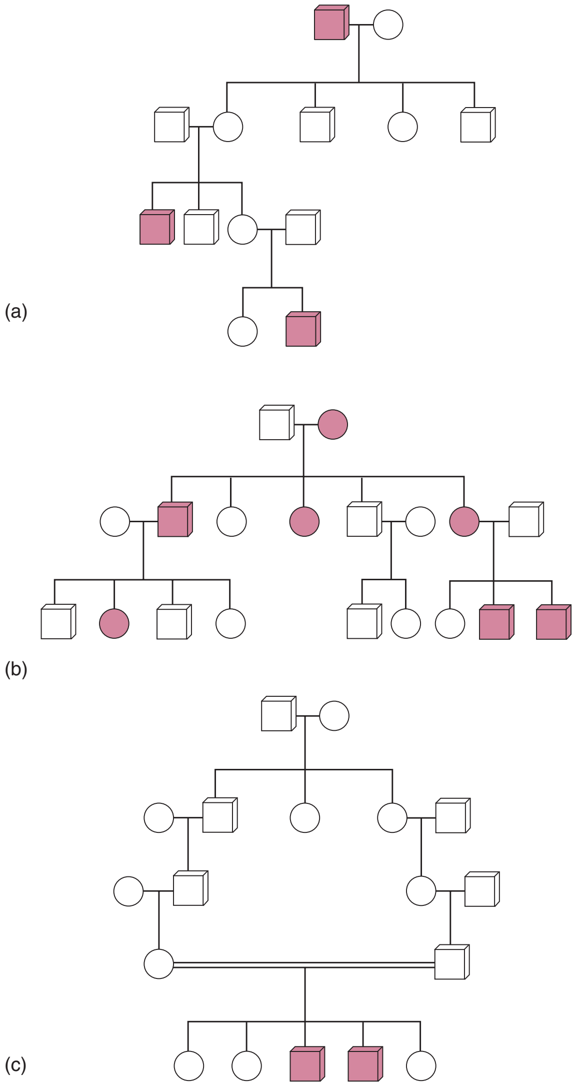
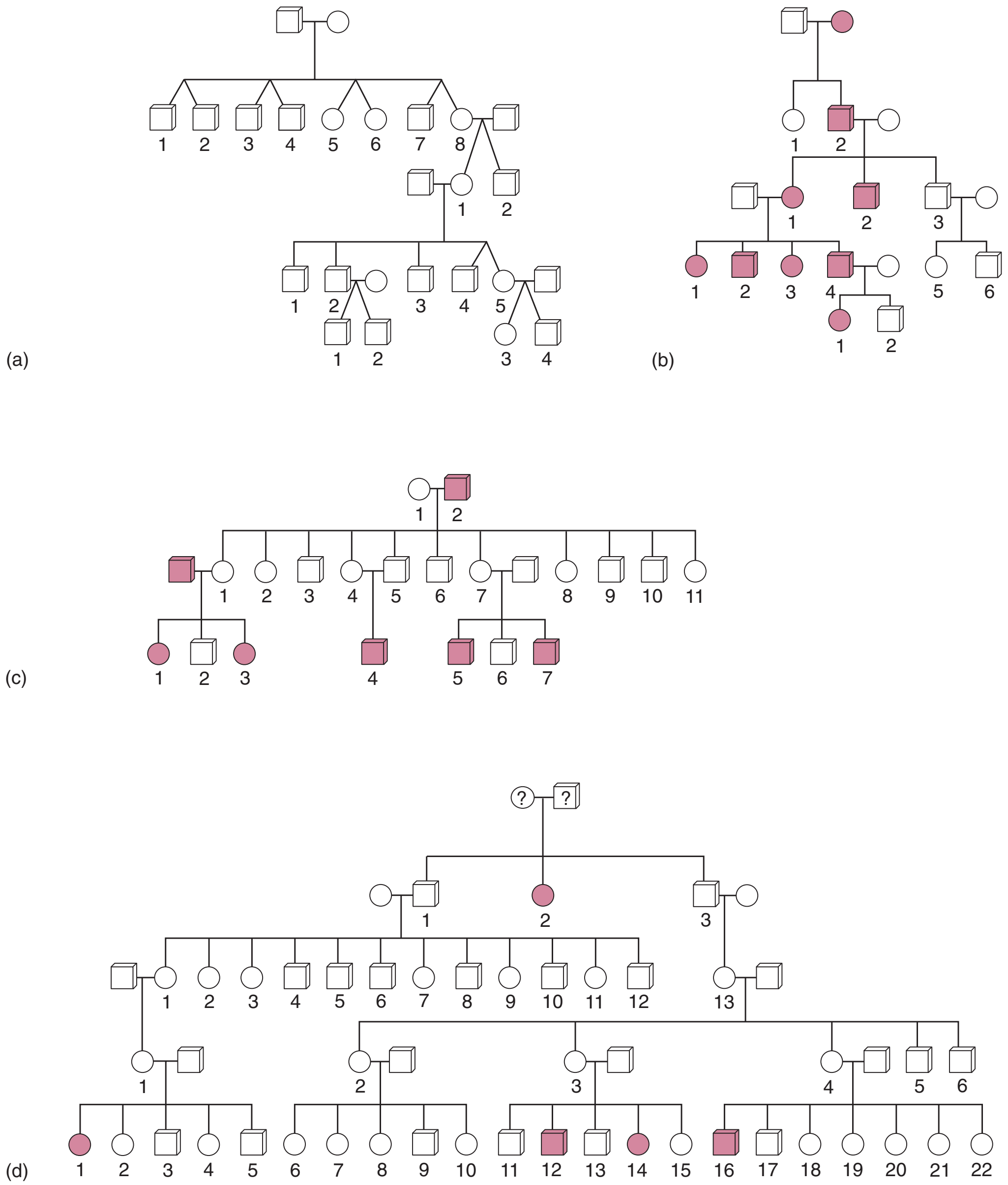


**Figure 4** Three pedigrees showing different modes of inheritance.

**25.** In pedigrees *a–d* in figure 5, which show the inheritance of rare human traits, including twin production, determine which modes of inheritance are most probable, possible, or impossible.

**26.** Hairy ears, a human trait expressed as excessive hair on the rims of ears in men, shows reduced penetrance (less than 100% penetrant). Mechanisms proposed include Y linkage, autosomal dominance, and autosomal recessiveness. Construct a pedigree consistent with each of these mechanisms.

**27.** Construct pedigrees for traits that *could not be*
- **a.** autosomal recessive.
- **b.** autosomal dominant.
- **c.** sex-linked recessive.
- **d.** sex-linked dominant.



(a)
1 2 3 4 5 6 7 8
1 2
1 2 3 4 5
1 2 3 4
(b)
1 2
1 2 3
1 2 3 4 5 6
1 2
(c)
1 2
1 2 3 4 5 6 7 8 9 10 11
1 2 3 4 5 6 7
(d)
?
?
1 2 3
1 2 3 4 5 6 7 8 9 10 11 12 13
1 2 3 4 5 6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

**Figure 5** Pedigrees of rare human traits, including twin production (*a*).

**28.** Determine the possible modes of inheritance for each trait in pedigrees *a–c* in figure 6.

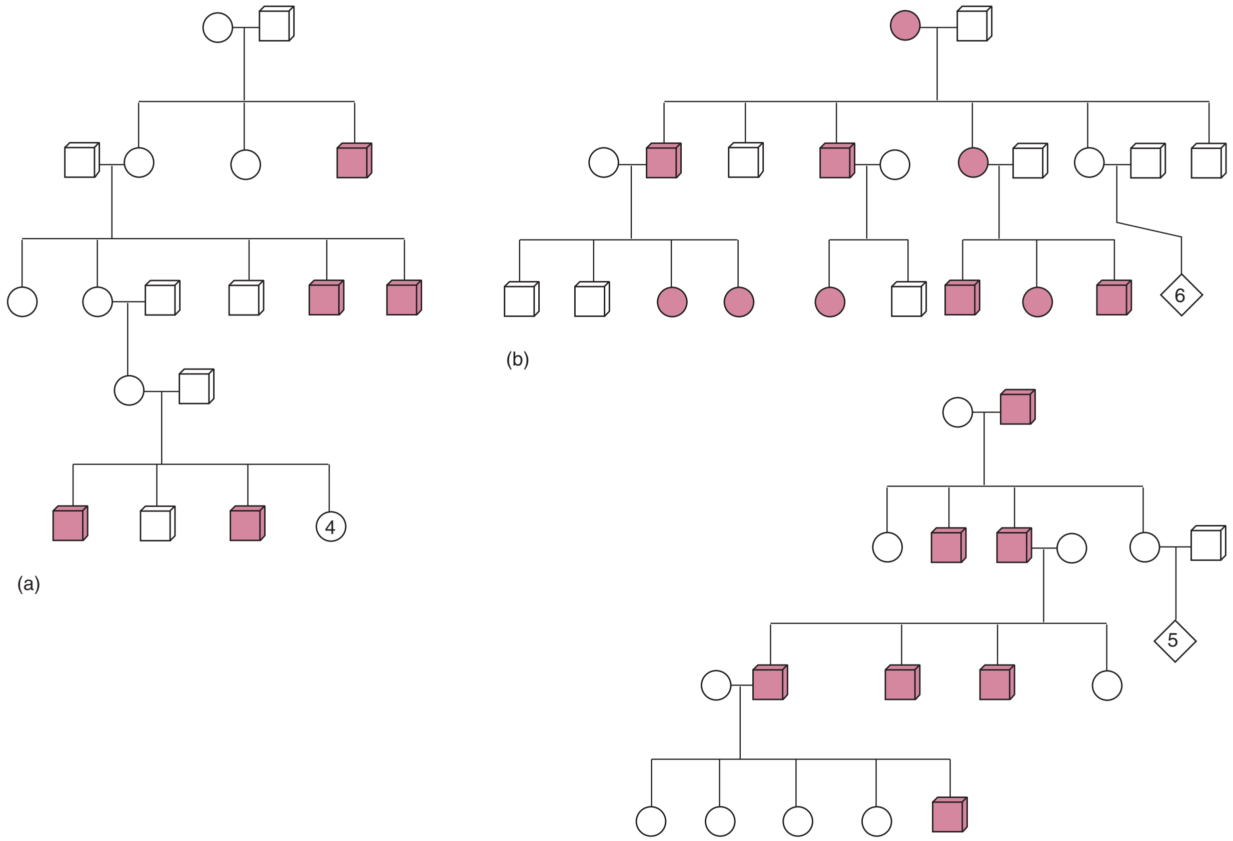


**Figure 6** Varying modes of inheritance.

## CRITICAL THINKING QUESTIONS

1. What effects do null alleles, alleles that produce no protein product, have in electrophoretic systems? How could you tell if a null allele were present?
2. In 1918, the Bolsheviks killed Tsar Nicholas II of Russia and his family (fig. 5.22). However, the remains of one daughter, Princess Anastasia, were never recovered. At one point, a woman appeared who claimed to be Anastasia. How could you validate her claim genetically?

*Suggested Readings for chapter 5 are on page B-2.*

# 6

# LINKAGE AND MAPPING IN EUKARYOTES

## STUDY OBJECTIVES

1. To learn about analytical techniques for locating the relative positions of genes on chromosomes in diploid eukaryotic organisms 110
2. To learn about analytical techniques for locating the relative positions of genes on chromosomes in ascomycete fungi 122
3. To learn about analytical techniques for locating the relative positions of genes on human chromosomes 132

## STUDY OUTLINE

**Diploid Mapping** 110
  Two-Point Cross 110
  Three-Point Cross 114
  Cytological Demonstration of Crossing Over 120
**Haploid Mapping (Tetrad Analysis)** 122
  Phenotypes of Fungi 124
  Unordered Spores (Yeast) 124
  Ordered Spores (*Neurospora*) 125
**Somatic (Mitotic) Crossing Over** 132
**Human Chromosomal Maps** 132
  X Linkage 132
  Autosomal Linkage 134
*Summary* 140
*Solved Problems* 140
*Exercises and Problems* 142
*Critical Thinking Questions* 147
**Box 6.1 The Nobel Prize** 112
**Box 6.2 The First Chromosomal Map** 121
**Box 6.3 *Lod* Scores** 135

Scanning electron micrograph (false color) of a fruit fly, *Drosophila melanogaster.*

(© Dr. Jeremy Burgess/SPL/Photo Researchers.)

After Sutton suggested the chromosomal theory of inheritance in 1903, evidence accumulated that genes were located on chromosomes. For example, Morgan showed by an analysis of inheritance patterns that the white-eye locus in Drosophila is located on the X chromosome. Given that any organism has many more genes than chromosomes, it follows that each chromosome has many loci. Since chromosomes in eukaryotes are linear, it also follows that genes are arranged in a linear fashion on chromosomes, like beads on a string. Sturtevant first demonstrated this in 1913. In this chapter, we look at analytical techniques for mapping chromosomes—techniques for determining the relationship between different genes on the same chromosome. These techniques are powerful tools that allow us to find out about the physical relationships of genes on chromosomes without ever having to see a gene or a chromosome. We determine that genes are on the same chromosome when the genes fail to undergo independent assortment, and then we use recombination frequencies to determine the distance between genes.

If loci were locked together permanently on a chromosome, allelic combinations would always be the same. However, at meiosis, crossing over allows the alleles of associated loci to show some measure of independence. A geneticist can use crossing over between loci to determine how close one locus actually is to another on a chromosome and thus to map an entire chromosome and eventually the entire **genome** (genetic complement) of an organism.

Loci carried on the same chromosome are said to be linked to each other. There are as many **linkage groups** ($l$) as there are autosomes in the haploid set plus sex chromosomes. *Drosophila* has five linkage groups ($2n = 8$; $l = 3$ autosomes + X + Y), whereas human beings have twenty-four linkage groups ($2n = 46$; $l = 22$ autosomes + X + Y). Prokaryotes and viruses, which usually have a single chromosome, are discussed in chapter 7.

Historically, classical mapping techniques, as described in this chapter and the next, gave researchers their only tools to determine the relationships of particular genes and their chromosomes. When geneticists know the locations of specific genes, they can study them in relation to each other and begin to develop a comprehensive catalogue of the genome of an organism. Knowing the location of a gene also helps in isolating the gene and studying its function and structure. And mapping the genes of different types of organisms (diploid, haploid, eukaryotic, prokaryotic) gives geneticists insight into genetic processes. More recently, recombinant DNA technology has allowed researchers to sequence whole genomes, including the human and fruit fly genomes; this means they now know the exact locations of all the genes on all the chromosomes of these organisms (see chapter 13). Geneticists are now creating massive databases containing this information, much of it available for free or by subscription on the World Wide Web. Until investigators mine all this information for all organisms of interest, they will still use analytical techniques in the laboratory and field to locate genes on chromosomes.

## DIPLOID MAPPING

### Two-Point Cross

In *Drosophila,* the recessive band gene (*bn*) causes a dark transverse band on the thorax, and the detached gene (*det*) causes the crossveins of the wings to be either detached or absent (fig. 6.1). A banded fly was crossed with a detached fly to produce wild-type, dihybrid offspring in the $F_1$ generation. $F_1$ females were then testcrossed to banded, detached males (fig. 6.2). (There is no crossing over in male fruit flies; in experiments designed to detect linkage, heterozygous females—in which crossing over occurs—are usually crossed with homozygous recessive males.) If the loci were assorting independently, we would expect a 1:1:1:1 ratio of the four possible phenotypes. However, of the first one thousand offspring examined, experimenters recorded a ratio of 2:483: 512:3.

Several points emerge from the data in figure 6.2. First, no simple ratio is apparent. If we divide by two, we get a ratio of 1:241:256:1.5. Although the first and last categories seem about equal, as do the middle two, no simple numerical relation seems to exist between the middle and end categories. Second, the two cate-

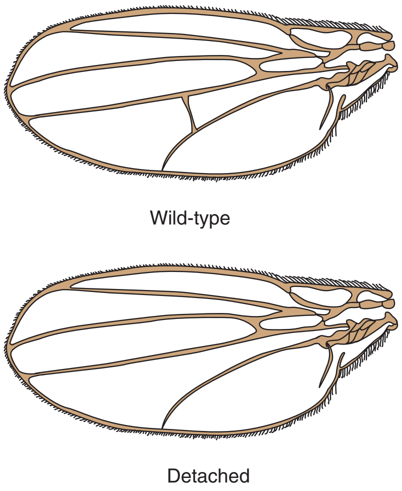


**Figure 6.1** Wild-type ($det^+$) and detached ($det$) crossveins in *Drosophila*.

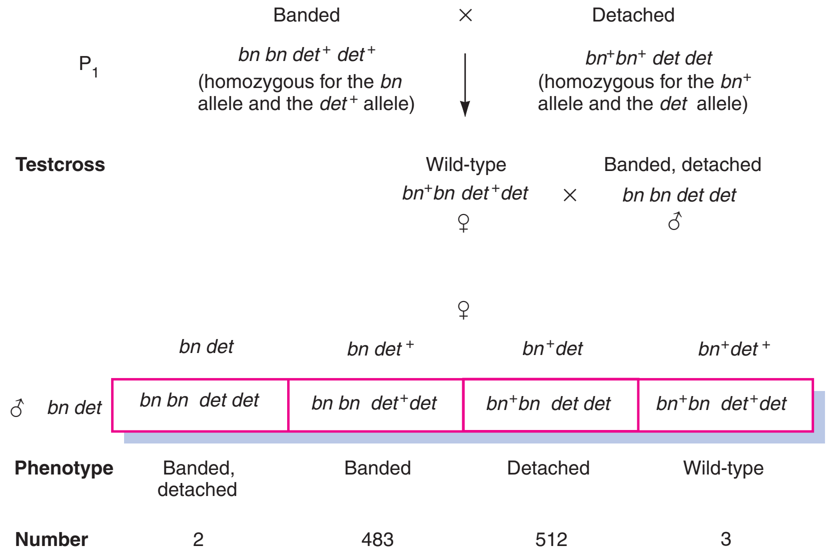


**Figure 6.2** Testcrossing a dihybrid *Drosophila.*

gories in very high frequency have the same phenotypes as the original parents in the cross ($P_1$ of fig. 6.2). That is, banded flies and detached flies were the original parents as well as the great majority of the testcross offspring. We call these phenotypic categories **parentals,** or **nonrecombinants.** On the other hand, the testcross offspring in low frequency combine the phenotypes of the two original parents ($P_1$). These two categories are referred to as **nonparentals,** or **recombinants.** The simplest explanation for these results is that the banded and detached loci are located near each other on the same chromosome (they are a linkage group), and therefore they move together as associated alleles during meiosis.

We can analyze the original cross by drawing the loci as points on a chromosome (fig. 6.3). This shows that 99.5% of the testcross offspring (the nonrecombinants) come about through the simple linkage of the two loci. The remaining 0.5% (the recombinants) must have arisen through a crossover of homologues, from a chiasma at meiosis, between the two loci (fig. 6.4). Note that since it is not possible to tell from these crosses which chromosome the loci are actually on or where the centromere is in relation to the loci, the centromeres are not included in the figures. The crossover event is viewed as a breakage and reunion of two chromatids lying adjacent to each other during prophase I of meiosis. Later in this chapter, we find cytological proof for this; in chapter 12, we explore the molecular mechanisms of this breakage and reunion process.

From the testcross in figure 6.3, we see that 99.5% of the gametes produced by the dihybrid are nonrecombinant, whereas only 0.5% are recombinant. This very small frequency of recombinant offspring indicates that the two loci lie very close to each other on their particular chromosome. In fact, we can use the recombination percentages of gametes, and therefore of testcross offspring, as estimates of distance between loci on a chromosome: 1% recombinant offspring is referred to as one **map unit** (or one **centimorgan,** in honor of geneticist T. H. Morgan, the first geneticist to win the Nobel Prize; box 6.1). Although a map unit is not a physical distance, it is a relative distance that makes it possible to know the order of and relative separation between loci on a chromosome. In this case, the two loci are 0.5 map units apart. (From sequencing various chromosomal segments—see chapter 13—we have learned that the relationship between centimorgans and DNA base pairs is highly variable, depending on species, sex, and region of the chromosome. For example, in human beings, 1 centimorgan can vary between 100,000 and 10,000,000 base pairs. In the fission yeast, *Schizosaccharomyces pombe,* 1 centimorgan is only about 6,000 base pairs.)

The arrangement of the *bn* and *det* alleles in the dihybrid of figure 6.3 is termed the **trans** configuration, meaning "across," because the two mutants are across from each other, as are the two wild-type alleles. The alternative arrangement, in which one chromosome carries both mutants and the other chromosome carries both wild-type alleles (fig. 6.5), is referred to as the **cis** configuration. (Two other terms, **repulsion** and **coupling,** have the same meanings as *trans* and *cis,* respectively.)

A cross involving two loci is usually referred to as a **two-point cross;** it gives us a powerful tool for dissecting the makeup of a chromosome. The next step in our

## BOX 6.1

## Historical Perspectives

### *The Nobel Prize*

On 10 December each year, the king of Sweden awards the Nobel Prizes at the Stockholm Concert Hall. The date is the anniversary of Alfred Nobel's death. Awards are given annually in physics, chemistry, medicine and physiology, literature, economics, and peace. In 2000, each award was worth $900,000, although an award sometimes goes to two or three recipients. The prestige is priceless.

Winners of the Nobel Prize are chosen according to the will of Alfred Nobel, a wealthy Swedish inventor and industrialist, who held over three hundred patents when he died in 1896 at the age of sixty-three. Nobel developed a detonator and processes for detonation of nitroglycerine, a substance invented by Italian chemist Ascanio Sobrero in 1847. In the form Nobel developed, the explosive was patented as dynamite. Nobel also invented several other forms of explosives. He was a benefactor of Sobrero, hiring him as a consultant and paying his wife a pension after Sobrero died.

Nobel believed that dynamite would be so destructive that it would serve as a deterrent to war. Later, realizing that this would not come to pass, he instructed that his fortune be invested and the interest used to fund the awards. The first prizes were awarded in 1901. Each award consists of a diploma, medal, and check.

American, British, German, French, and Swedish citizens have earned the most prizes (table 1). Table 2 features some highlights of Nobel laureate achievements in genetics.

ALFR·
NOBEL
NAT·
MDCCC
XXXIII
OB·
MDCCC
XCVI

**The Nobel medal.** The medal is half a pound of 23-karat gold, measures about 2 1/2 inches across, and has Nobel's face and the dates of his birth and death on the front. The diplomas that accompany the awards are individually designed. (Reproduced by permission of the Nobel Foundation.)

**Table 1** **Distribution of Nobel Awards to the Top Five Recipient Nations (Including 2000 Winners)**

| | Physics | Chemistry | Medicine and Physiology | Peace | Literature | Economics | Total |
|---|---|---|---|---|---|---|---|
| United States | 77 | 46 | 88 | 20 | 9 | 30 | 270 |
| Britain | 20 | 24 | 25 | 9 | 8 | 5 | 91 |
| Germany | 18 | 27 | 15 | 4 | 6 | 1 | 71 |
| France | 12 | 7 | 7 | 8 | 12 | 1 | 47 |
| Sweden | 4 | 4 | 7 | 5 | 7 | 2 | 29 |

**Table 2** **Some Nobel Laureates in Genetics (Medicine and Physiology; Chemistry)**

| Name | Year | Nationality | Cited for |
|---|---|---|---|
| Thomas Hunt Morgan | 1933 | USA | Discovery of how chromosomes govern heredity |
| Hermann J. Muller | 1946 | USA | X-ray inducement of mutations |
| George W. Beadle | 1958 | USA | Genetic regulation of biosynthetic pathways |
| Edward L. Tatum | 1958 | USA | |
| Joshua Lederberg | 1958 | USA | Bacterial genetics |
| Severo Ochoa | 1959 | USA | Discovery of enzymes that synthesize nucleic acids |
| Arthur Kornberg | 1959 | USA | |
| Francis H. C. Crick | 1962 | British | Discovery of the structure of DNA |
| James D. Watson | 1962 | USA | |
| Maurice Wilkins | 1962 | British | |
| François Jacob | 1965 | French | Regulation of enzyme biosynthesis |
| André Lwoff | 1965 | French | |
| Jacques Monod | 1965 | French | |
| Peyton Rous | 1966 | USA | Tumor viruses |
| Robert W. Holley | 1968 | USA | Unraveling of the genetic code |
| H. Gobind Khorana | 1968 | USA | |
| Marshall W. Nirenberg | 1968 | USA | |
| Max Delbrück | 1969 | USA | Viral genetics |
| Alfred Hershey | 1969 | USA | |
| Salvador Luria | 1969 | USA | |
| Renato Dulbecco | 1975 | USA | Tumor viruses |
| Howard Temin | 1975 | USA | Discovery of reverse transcriptase |
| David Baltimore | 1975 | USA | |
| Werner Arber | 1978 | Swiss | Discovery and use of restriction endonucleases |
| Hamilton Smith | 1978 | USA | |
| Daniel Nathans | 1978 | USA | |
| Walter Gilbert | 1980 | USA | Techniques of sequencing DNA |
| Frederick Sanger | 1980 | British | |
| Paul Berg | 1980 | USA | Pioneer work in recombinant DNA |
| Baruj Benacerraf | 1980 | USA | Genetics of immune reactions |
| Jean Dausset | 1980 | French | |
| George Snell | 1980 | USA | |
| Aaron Klug | 1982 | British | Crystallographic work on protein-nucleic acid complexes |
| Barbara McClintock | 1983 | USA | Transposable genetic elements |
| Cesar Milstein | 1984 | British/Argentine | Immunogenetics |
| Georges Koehler | 1984 | German | |
| Niels K. Jerne | 1984 | British/Danish | |
| Susumu Tonegawa | 1987 | Japanese | Antibody diversity |
| J. Michael Bishop | 1989 | USA | Proto-oncogenes |
| Harold E. Varmus | 1989 | USA | |
| Thomas R. Cech | 1989 | USA | Enzymatic properties of RNA |
| Sidney Altman | 1989 | Canada | |
| Kary Mullis | 1993 | USA | Polymerase chain reaction |
| Michael Smith | 1993 | Canada | Site-directed mutagenesis |
| Richard Roberts | 1993 | British | Discovery of intervening sequences in RNA |
| Phillip Sharp | 1993 | USA | |
| E. B. Lewis | 1995 | USA | Genes control development |

*continued*

## BOX 6.1 CONTINUED

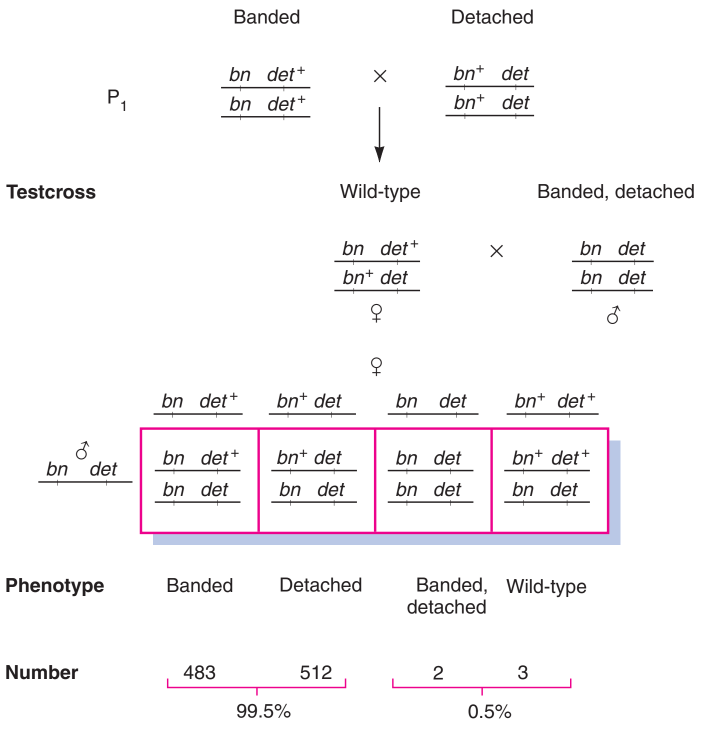
**Table 2** *continued*

| Name | Year | Nationality | Cited for |
|---|---|---|---|
| Christiane Nüsslein-Volhard | 1995 | German | |
| Eric Wieschaus | 1995 | USA | |
| Stanley B. Prusiner | 1997 | USA | Discovery of prions |
| Günter Blobel | 1999 | German | Signal recognition during protein synthesis |

analysis is to look at three loci simultaneously so that we can determine their relative order on the chromosome. More important, we can also analyze the effects of multiple crossovers, which cannot be detected in a two-point cross, on map distances. Two crossovers between two loci can cause the chromosome to look as if no crossovers took place, causing us to underestimate map distances. Thus we need a third locus, between the first two, to detect multiple crossover events.

### *Three-Point Cross*

Analysis of three loci, each segregating two alleles, is referred to as a **three-point cross.** We will examine wing morphology, body color, and eye color in *Drosophila.* Black body (*b*), purple eyes (*pr*), and curved wings (*c*) are all recessive genes. Since the most efficient way to study linkage is through the testcross of a multihybrid, we will study these three loci by means of the crosses shown in



Banded
Detached
P1
bn det+
bn det+
×
bn+ det
bn+ det
Testcross
Wild-type
Banded, detached
bn det+
bn+ det
♀
×
bn det
bn det
♂
♀
bn det+
bn+ det
bn det
bn+ det+
♂
bn det
bn det+
bn det
bn+ det
bn det
bn det
bn det
bn+ det+
bn det
Phenotype
Banded
Detached
Banded, detached
Wild-type
Number
483
512
2
3
99.5%
0.5%

**Figure 6.3** Chromosomal arrangement of the two loci in the crosses of figure 6.2. A line arbitrarily represents the chromosomes on which these loci are actually situated.

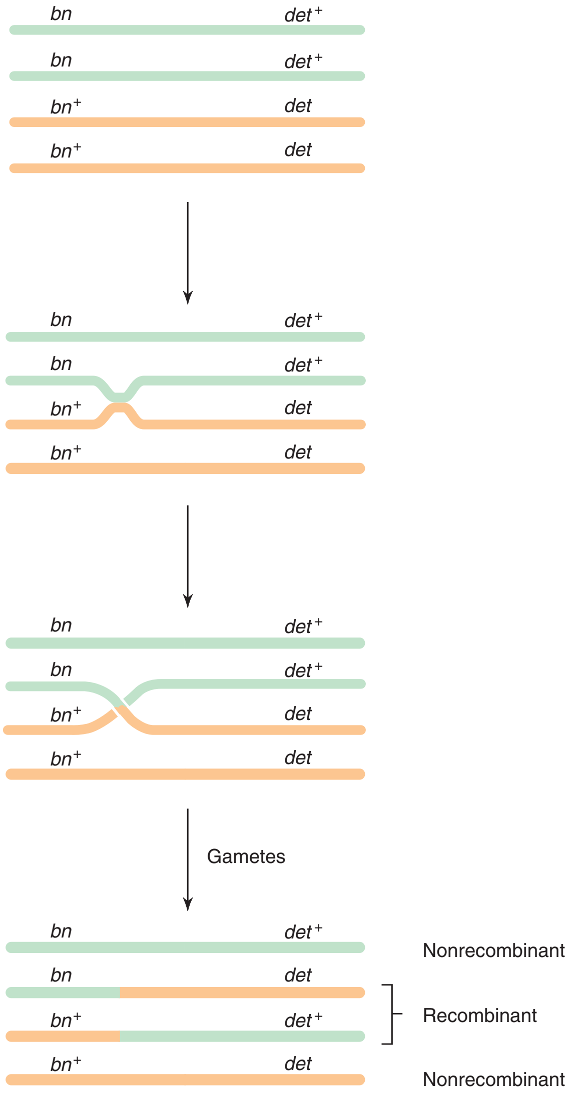


**Figure 6.4** Crossover of homologues during meiosis between the *bn* and *det* loci in the tetrad of the dihybrid female.

figure 6.6. One point in this figure should be clarified. Since the organisms are diploid, they have two alleles at each locus. Geneticists use various means to present this situation. For example, the recessive homozygote can be pictured as

1. *bb prpr cc*
2. *b/b pr/pr c/c* or $\frac{b}{b}\ \frac{pr}{pr}\ \frac{c}{c}$
3. *b pr c/b pr c* or $\frac{b\ \ pr\ \ c}{b\ \ pr\ \ c}$

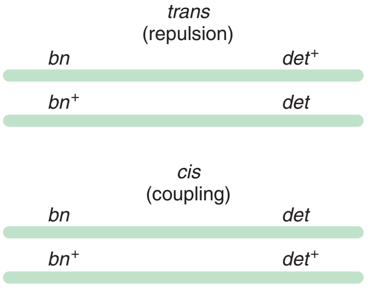
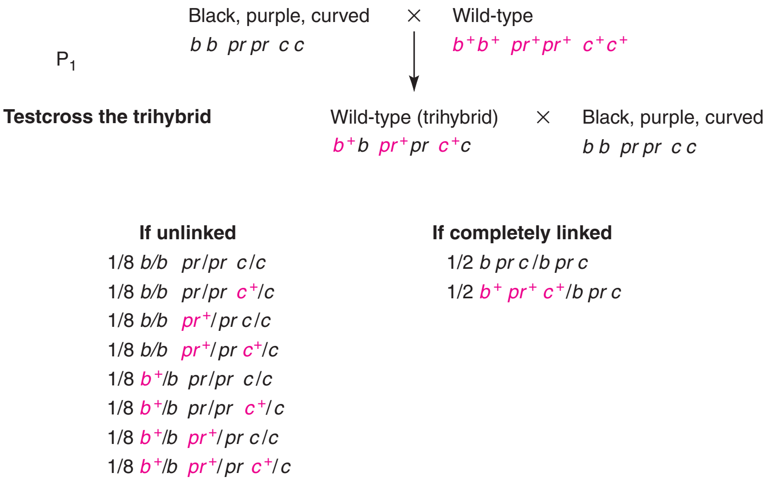


**Figure 6.5** *Trans* (repulsion) and *cis* (coupling) arrangements of dihybrid chromosomes.

A slash (also called a rule line) is used to separate alleles on homologous chromosomes. Thus (*1*) is used tentatively, when we do not know the linkage arrangement of the loci, (*2*) is used to indicate that the three loci are on different chromosomes, and (*3*) indicates that all three loci are on the same chromosome.

In figure 6.6, the trihybrid organism is testcrossed. If independent assortment is at work, the eight types of resulting gametes should appear with equal frequencies, and thus the eight phenotypic classes would each make up one-eighth of the offspring. However, if there were *complete linkage,* so that the loci are so close together on the same chromosome that virtually no crossing over takes place, we would expect the trihybrid to produce only two gamete types in equal frequency and to yield two phenotypic classes identical to the original parents. This would occur because, under complete linkage, the trihybrid would produce only two chromosomal types in gametes: the *b pr c* type from one parent and the $b^+$ $pr^+$ $c^+$ type from the other. Crossing over between linked loci would produce eight phenotypic classes in various proportions depending on the distances between loci. The actual data appear in table 6.1.

The data in the table are arranged in reciprocal classes. Two classes are reciprocal if between them they contain each mutant phenotype just once. Wild-type and black, purple, curved classes are thus reciprocal, as are the purple, curved and the black classes. Reciprocal classes occur in roughly equal numbers: 5,701 and 5,617; 388 and 367; 1,412 and 1,383; and 60 and 72. As we shall see, a single meiotic recombinational event produces reciprocal classes. Wild-type and black, purple, curved are the two nonrecombinant classes. The purple, curved class of 388 is grouped with the black class of 367. These two would be the products of a crossover between the *b* and the *pr* loci if we assume that the three loci are linked and that the gene order is *b pr c* (fig. 6.7). The next two classes, of 1,412 and 1,383 flies, would result from a crossover between *pr* and *c,* and the last set, 60 and 72, would result from two crossovers, one between *b* and *pr* and the other between



Black, purple, curved
b b pr pr c c
×
Wild-type
b⁺b⁺ pr⁺pr⁺ c⁺c⁺
P1
Testcross the trihybrid
Wild-type (trihybrid)
b⁺b pr⁺pr c⁺c
×
Black, purple, curved
b b pr pr c c
If unlinked
1/8 b/b pr/pr c/c
1/8 b/b pr/pr c⁺/c
1/8 b/b pr⁺/pr c/c
1/8 b/b pr⁺/pr c⁺/c
1/8 b⁺/b pr/pr c/c
1/8 b⁺/b pr/pr c⁺/c
1/8 b⁺/b pr⁺/pr c/c
1/8 b⁺/b pr⁺/pr c⁺/c
If completely linked
1/2 b pr c/b pr c
1/2 b⁺ pr⁺ c⁺/b pr c

**Figure 6.6** Possible results in the testcross progeny of the *b pr c* trihybrid.

*pr* and *c* (fig. 6.8). Groupings according to these recombinant events are shown at the right in table 6.1.

In the final column of table 6.1, recombination between *b* and *c* is scored. Only those recombinant classes that have a new arrangement of *b* and *c* alleles, as compared with the parentals, are counted. This last column shows us what a *b–c,* two-point cross would have revealed had we been unaware of the *pr* locus in the middle.

### *Map Distances*

The percent row in table 6.1 reveals that 5.9% (887/15,000) of the offspring in the *Drosophila* trihybrid testcross resulted from recombination between *b* and *pr;* 19.5% between *pr* and *c,* and 23.7% between *b* and *c.* These numbers allow us to form a tentative map of the loci (fig. 6.9). There is, however, a discrepancy. The distance between *b* and *c* can be calculated in two ways. By adding the two distances, *b–pr* and *pr–c,* we get 5.9 + 19.5 = 25.4 map units; yet by directly counting the recombinants (the last column of table 6.1), we get a distance of only 23.7 map units. What causes this discrepancy of 1.7 map units?

Returning to the last column of table 6.1, we observe that the double crossovers (60 and 72) are not counted, yet each actually represents two crossovers in this re-

**Table 6.1** **Results of Testcrossing Female *Drosophila* Heterozygous for Black Body Color, Purple Eye Color, and Curved Wings ($b^+b\ pr^+pr\ c^+c \times bb\ prpr\ cc$)**

| Phenotype | Genotype | Number | Alleles from Trihybrid Female | Number Recombinant Between | | |
|---|---|---|---|---|---|---|
| | | | | ***b* and *pr*** | ***pr* and *c*** | ***b* and *c*** |
| Wild-type | $b^+b\ pr^+pr\ c^+\ c$ | 5,701 | $b^+\ pr^+\ c^+$ | | | |
| Black, purple, curved | *bb prpr cc* | 5,617 | *b pr c* | | | |
| Purple, curved | $b^+b\ prpr\ cc$ | 388 | $b^+\ pr\ c$ | 388 | | 388 |
| Black | $bb\ pr^+pr\ c^+c$ | 367 | $b\ pr^+\ c^+$ | 367 | | 367 |
| Curved | $b^+b\ pr^+pr\ cc$ | 1,412 | $b^+\ pr^+\ c$ | | 1,412 | 1,412 |
| Black, purple | $bb\ prpr\ c^+c$ | 1,383 | $b\ pr\ c^+$ | | 1,383 | 1,383 |
| Purple | $b^+b\ prpr\ c^+c$ | 60 | $b^+pr\ c^+$ | 60 | 60 | |
| Black, curved | $bb\ pr^+pr\ cc$ | 72 | $b\ pr^+\ c$ | 72 | 72 | |
| Total | | 15,000 | | 887 | 2,927 | 3,550 |
| Percent | | | | 5.9 | 19.5 | 23.7 |

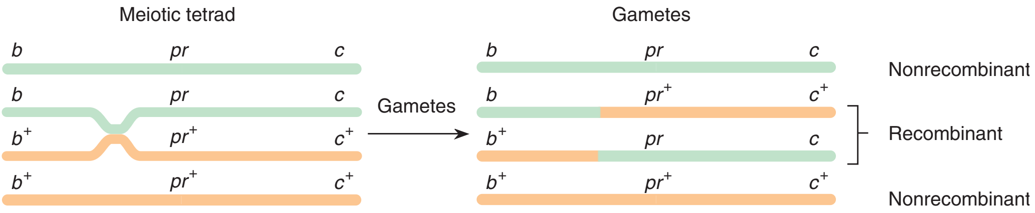

Meiotic tetrad
Gametes
b
pr
c
b
pr
c
$b^{+}$
$pr^{+}$
$c^{+}$
$b^{+}$
$pr^{+}$
$c^{+}$
Gametes
b
pr
c
b
$pr^{+}$
$c^{+}$
$b^{+}$
pr
c
$b^{+}$
$pr^{+}$
$c^{+}$
Nonrecombinant
Recombinant
Nonrecombinant

**Figure 6.7** Results of a crossover between the black and purple loci in *Drosophila.*

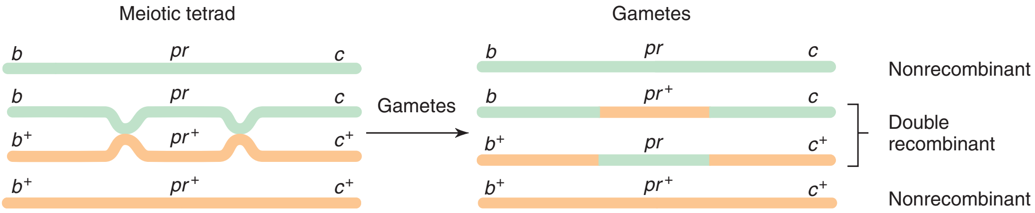

Meiotic tetrad
Gametes
b
pr
c
b
pr
c
$b^{+}$
$pr^{+}$
$c^{+}$
$b^{+}$
$pr^{+}$
$c^{+}$
Gametes
b
pr
c
b
$pr^{+}$
c
$b^{+}$
pr
$c^{+}$
$b^{+}$
$pr^{+}$
$c^{+}$
Nonrecombinant
Double recombinant
Nonrecombinant

**Figure 6.8** Results of a double crossover in the *b pr c region of the Drosophila* chromosome.

gion. The reason they are not counted is simply that if we observed only the end loci of this chromosomal segment, we would not detect the double crossovers; the first one of the two crossovers causes a recombination between the two end loci, whereas the second one returns these outer loci to their original configuration (see fig. 6.8). If we took the 3,550 recombinants between *b* and *c* and added in twice the total of the double recombinants, 264, we would get a total of 3,814. This is 25.4 map units, which is the more precise figure we calculated before. The farther two loci are apart on a chromosome, the more double crossovers occur between them. Double crossovers tend to mask recombinants, as in our example, so that distantly linked loci usually appear closer than they really are. Thus, the most accurate map distances are those established on very closely linked loci. In other words, summed short distances are more accurate than directly measured larger distances.

The results of the previous experiment show that we can obtain at least two map distances between any two loci: measured and actual. Measured map distance between two loci is the value obtained from a two-point cross. Actual map distance is an idealized, more accurate value obtained from summing short distances between many intervening loci. We obtain the short distances from crosses involving other loci between the original two. When we plot measured map distance against actual map distance, we obtain the curve in figure 6.10. This curve is called a **mapping function.** This graph is of both practical and theoretical value. Pragmatically, it allows

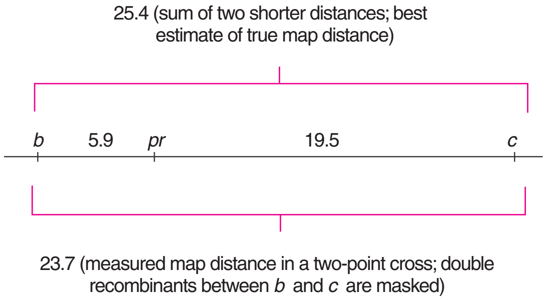

25.4 (sum of two shorter distances; best estimate of true map distance)
b
5.9
pr
19.5
c
23.7 (measured map distance in a two-point cross; double recombinants between b and c are masked)

**Figure 6.9** Tentative map of the black, purple, and curved chromosome in *Drosophila.* Numbers are map units (centimorgans).

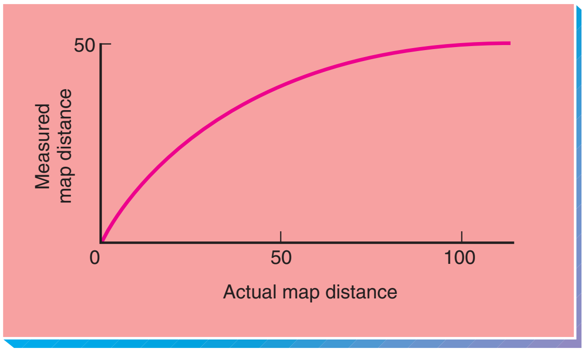


**Figure 6.10** Mapping function.

us to convert a measured map distance into a more accurate one. Theoretically, it shows that measured map distance never exceeds 50 map units in any one cross. Multiple crossovers reduce the apparent distance between two loci to a maximum of 50 map units, the value that independent assortment produces (50% parentals, 50% recombinants).

### Gene Order

Although we performed the previous analysis merely assuming that *pr* was in the middle, the data in table 6.1 confirm our original assumption that the gene order is *b pr c.* Of the four pairs of reciprocal phenotypic classes in table 6.1, one pair has the highest frequency (5,701 and 5,617) and one pair has the lowest (60 and 72). The pair with highest frequency is the nonrecombinant group. The one with the lowest frequency is the double recombinant group, the one in which only the middle locus has been changed from the parental arrangement. A comparison of either of the double recombinant classes with either of the nonrecombinant classes shows the gene that is in the middle and, therefore, the gene order. In other words, the data allow us to determine gene order. Since $b^+$ $pr^+$ $c^+$ was one of the nonrecombinant gametes, and $b^+$ *pr* $c^+$ was one of the double recombinant gametes, *pr* stands out as the changed locus, or the one in the middle. In a similar manner, comparing *b* $pr^+$ *c* with *b pr c* would also point to *pr* as the inside locus (or **inside marker**). So would comparing $b^+$ *pr* $c^+$ with *b pr c* or *b* $pr^+$ *c* with $b^+$ *pr* $c^+$. In each case, the middle locus, *pr,* displays the different pattern, whereas the allelic arrangements of the outside markers, *b* and *c,* behave in concert.

If this seems confusing, simply compare the double crossovers and nonrecombinants to find one of each in which two alleles are identical. For example, the double recombinant $b^+$ *pr* $c^+$ and the nonrecombinant $b^+$ $pr^+$ $c^+$ share the $b^+$ and $c^+$ alleles. The *pr* locus is mutant in one case and wild-type in the other. Hence, *pr* is the locus in the middle.

From the data in table 6.1, we can confirm the association of alleles in the trihybrid parent. That is, since the data came from testcrossing a trihybrid, the allelic configuration in that trihybrid is reflected in the nonrecombinant classes of offspring. In this case, one is the result of a $b^+$ $pr^+$ $c^+$ gamete, the other, of a *b pr c* gamete. Thus, the trihybrid had the genotype *b pr c*/$b^+$ $pr^+$ $c^+$: all alleles were in the *cis* configuration.

### Coefficient of Coincidence

The next question in our analysis of this three-point cross is, are crossovers occurring independently of each other? That is, does the observed number of double recombinants equal the expected number? In the example, there were 132/15,000 double crossovers, or 0.88%. The expected number is based on the independent occurrence of crossing over in the two regions measured. That is, 5.9% of the time there is a crossover in the *b–pr* region, which we can express as a probability of occurrence of 0.059. Similarly, 19.5% of the time there is a crossover in the *pr–c* region, or a probability of occurrence of 0.195. A double crossover should occur as a product of the two probabilities: $0.059 \times 0.195 = 0.0115$. This means that 1.15% of the gametes (1.15% of 15,000 = 172.5) should be double recombinants. In our example, the observed number of double recombinant offspring is lower than expected (132 observed, 172.5 expected). This implies a **positive interference,** in which the occurrence of the first crossover reduces the chance of the second. We can express this as a **coefficient of coincidence,** defined as

$$\frac{\text{observed number of double recombinants}}{\text{expected number of double recombinants}}$$

In the example, the coefficient of coincidence is $132/172.5 = 0.77$. In other words, only 77% of the expected double crossovers occurred. Sometimes we express this reduced quantity of double crossovers as the *degree of interference,* defined as

$$\text{interference} = 1 - \text{coefficient of coincidence}$$

In our example, the interference is 23%.

It is also possible to have **negative interference,** in which we observe more double recombinants than expected. In this situation, the occurrence of one crossover seems to enhance the probability that crossovers will occur in adjacent regions.

### Another Example

Let us work out one more three-point cross, in which neither the middle gene nor the *cis-trans* relationship of the alleles in the trihybrid $F_1$ parent is given. On the third chromosome of *Drosophila,* hairy (*h*) causes extra bristles on the body, thread (*th*) causes a thread-shaped arista (antenna tip), and rosy (*ry*) causes the eyes to be reddish brown. All three traits are recessive. Trihybrid females were testcrossed; the phenotypes from one thousand offspring are listed in table 6.2. At this point, it is possible to use the data to determine the parental genotypes (the $P_1$ generation, assuming that they were homozygotes), the gene order, the map distances, and the coefficient of coincidence. The table presents the data in no particular order, as an experimenter might have recorded them. Phenotypes are tabulated and, from these, the genotypes can be reconstructed. Notice that the data can be put into the form found in table 6.1;

**Table 6.2** **Offspring from a Trihybrid ($h^+h\ ry^+ry\ th^+th$) Testcross (x *hh ryry thth*) in *Drosophila***

| Phenotype | Genotype (order unknown) | Number |
|---|---|---|
| Thread | $h^+ry^+th/h\ ry\ th$ | 359 |
| Rosy, thread | $h^+ry\ th/h\ ry\ th$ | 47 |
| Hairy, rosy, thread | $h\ ry\ th/h\ ry\ th$ | 4 |
| Hairy, thread | $h\ ry^+th/h\ ry\ th$ | 98 |
| Rosy | $h^+ry\ th^+/h\ ry\ th$ | 92 |
| Hairy, rosy | $h\ ry\ th^+/h\ ry\ th$ | 351 |
| Wild-type | $h^+ry^+th^+/h\ ry\ th$ | 6 |
| Hairy | $h\ ry^+th^+/h\ ry\ th$ | 43 |

**Table 6.3** **Data from Table 6.2 Arranged to Show Recombinant Regions**

| Trihybrid's Gamete | Number | *h–th* | *th–ry* | *h–ry* |
|---|---|---|---|---|
| $h^+th\ ry^+$ | 359 | | | |
| $h\ th^+ry$ | 351 | | | |
| $h\ th\ ry^+$ | 98 | 98 | | 98 |
| $h^+th^+ry$ | 92 | 92 | | 92 |
| $h^+th\ ry$ | 47 | | 47 | 47 |
| $h\ th^+\mathrm{ry}^+$ | 43 | | 43 | 43 |
| $h\ th\ ry$ | 4 | 4 | 4 | |
| $h^+th^+ry^+$ | 6 | 6 | 6 | |
| Total | 1,000 | 200 | 100 | 280 |

we see a large reciprocal set (359 and 351), a small reciprocal set (4 and 6), and large and small intermediate sets (98 and 92, 47 and 43).

From the data presented, is it obvious that the three loci are linked? The pattern, as just mentioned, is identical to that of the previous example, in which the three loci were linked. (What pattern would appear if two of the loci were linked and one assorted independently? See problem 6 at the end of the chapter.) Next, what is the allelic arrangement in the trihybrid parent? The offspring with the parental, or nonrecombinant, arrangements are the reciprocal pair in highest frequency. Table 6.2 shows that thread and hairy, rosy offspring are the nonrecombinants. Thus, the nonrecombinant gametes of the trihybrid $F_1$ parent were $h\ ry\ th^+$ and $h^+\ ry^+\ th$, which is the allelic arrangement of the trihybrid with the actual order still unknown—$h\ ry\ th^+/h^+\ ry^+\ th$. (What were the genotypes of the parents of this trihybrid, assuming they were homozygotes?) Continuing, which gene is in the middle? From table 6.2, we know that *h ry th* and $h^+\ ry^+\ th^+$ are the double recombinant gametes of the trihybrid parent because they occur in such low numbers. Comparison of these chromosomes with either of the nonrecombinant chromosomes ($h^+\ ry^+\ th$ or $h\ ry\ th^+$) shows that the thread (*th*) locus is in the middle. We now know that the original trihybrid had the following chromosomal composition: $h\ th^+\ ry/h^+\ th\ ry^+$. The *h* and *ry* alleles are in the *cis* configuration, with *th* in the *trans* configuration.

We can now compare the chromosome from the trihybrid in each of the eight offspring categories with the parental arrangement and determine the regions that had crossovers. Table 6.3 does this. We can see that the *h–th* distance is 20 map units, the *th–ry* distance is 10 map units, and the apparent *h–ry* distance is 28 map units (fig. 6.11). As in the earlier example, the *h–ry* discrepancy is from not counting the double crossovers twice each: 280 + 2(10) = 300, which is 30 map units and the more accurate figure. Last, we wish to know what the coefficient of coincidence is. The expected occurrence of double recombinants is $0.200 \times 0.100 = 0.020$, or 2%. Two percent of 1,000 = 20. Thus

$$\text{coefficient of coincidence} = \frac{\text{observed number of double recombinants}}{\text{expected number of double recombinants}} = 10/20 = 0.50$$

Only 50% of the expected double crossovers occurred.

Geneticists have mapped the chromosomes of many eukaryotic organisms from three-point crosses of this type—those of *Drosophila* are probably the most extensively studied. *Drosophila* and other species of flies have giant **polytene** salivary gland chromosomes, which arise as a result of **endomitosis.** In this process,

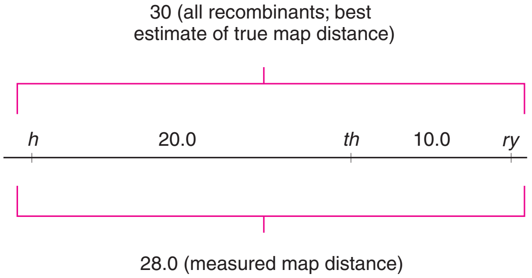

30 (all recombinants; best estimate of true map distance)
h
20.0
th
10.0
ry
28.0 (measured map distance)

**Figure 6.11** Map of the *h th ry* region of the *Drosophila* chromosome, with numerical discrepancy in distances. Numbers are map units (centimorgans).

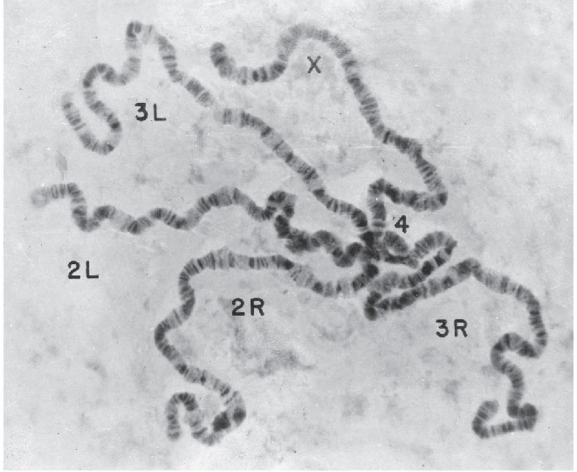


**Figure 6.12** Giant salivary gland chromosomes of *Drosophila.* X, 2, 3, and 4 are the four nonhomologous chromosomes. *L* and *R* indicate the *left* and *right* arms (in relation to the centromere). The dark bands are chromomeres. (B. P. Kaufman, "Induced Chromosome Rearrangements in *Drosophila melanogaster,*" *Journal of Heredity,* 30:178–90, 1939. Reproduced by permission of Oxford University Press.)

the chromosomes replicate, but the cell does not divide. In the salivary gland of the fruit fly, homologous chromosomes synapse and then replicate to make about one thousand copies, forming very thick structures with a distinctive pattern of bands called chromomeres (fig. 6.12). Using methods chapter 8 will discuss, scientists have mapped many loci to particular bands. Part of the *Drosophila* chromosomal map is presented in figure 6.13 (see also box 6.2). Locate the loci we have mapped so far to verify the map distances.

In summary, we know that two or more loci are linked if offspring do not fall into simple Mendelian ratios. Map distances are the percentage of recombinant offspring in a testcross. With three loci, determine the parental (nonrecombinant) and double recombinant groups first. Then establish the locus in the middle, and recast the data in the correct gene order. The most accurate map distances are those obtained by summing shorter distances. Determine a coefficient of coincidence by comparing observed number of double recombinants to expected number.

## *Cytological Demonstration of Crossing Over*

If we are correct that a chiasma during meiosis is the visible result of a physical crossover, then we should be able to demonstrate that genetic crossing over is accompanied by cytological crossing over. That is, the recombination event should entail the exchange of physical parts of homologous chromosomes. This can be demonstrated if we can distinguish between two homologous chromosomes, a technique Creighton and McClintock first used in maize (corn) and Stern first applied to *Drosophila,* both in 1931. We will look at Creighton and McClintock's experiment.

Harriet Creighton and Barbara McClintock worked with chromosome 9 in maize ($n = 10$). In one strain, they found a chromosome with abnormal ends. One end had a knob, and the other had an added piece of chromatin from another chromosome (fig. 6.14). This knobbed chromosome was thus clearly different from its normal homologue. It also carried the dominant colored (*C*) allele and the recessive waxy texture (*wx*) allele. After mapping studies showed that *C* was very close to the knob and *wx* was close to the added piece of chromatin, Creighton and McClintock made the cross shown in figure 6.14. The dihybrid plant with heteromorphic chromosomes was crossed with the normal homomorphic plant (only normal chromosomes) that had the genotype of *c Wx/c wx* (colorless and nonwaxy phenotype). If a crossover occurred during meiosis in the dihybrid in the region between *C* and *wx*, a physical crossover, visible cytologically (under the microscope), should also occur, causing the knob to become associated with an otherwise normal chromosome and the extra piece of chromosome 9 to be associated with a knobless chromosome. Four types of gametes would result (fig. 6.14).

Barbara McClintock (1902–1992). (Courtesy of Cold Spring Harbor Research Library Archives. Photographer, David Miklos.)

Harriet B. Creighton (1909– ). (Courtesy of Harriet B. Creighton.)

## BOX 6.2

## Historical Perspectives

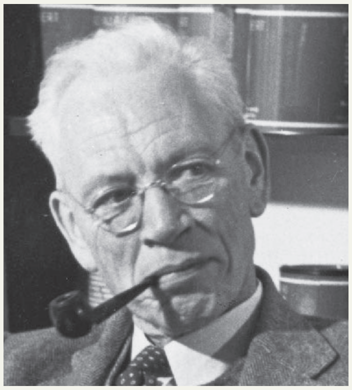
### *The First Chromosomal Map*

The first chromosomal map ever published included just five loci on the X chromosome of *Drosophila melanogaster* (fig. 1). It was published in 1913 by Alfred H. Sturtevant, who began working in Thomas Hunt Morgan's "fly lab" while an undergraduate student at Columbia University. The fly lab included H. J. Muller, later to win a Nobel Prize, and Calvin B. Bridges, whose work on sex determination in *Drosophila* we discussed in the last chapter.

Sturtevant worked with six mutants: yellow body ($y$); white ($w$), eosin ($w^e$), and vermilion eyes ($v$); and miniature ($m$) and rudimentary wings ($r$). (White and eosin are actually allelic; Sturtevant found no crossing over between the two "loci.") Using crosses similar to the ones we outline in this chapter, he constructed the map shown in figure 1. The map distances we accept today are very similar to the ones he obtained.

Sturtevant's work was especially important at this point because his data supported several basic concepts, including the linear arrangement of genes, which argued for the placement of genes on chromosomes as the only linear structures in the nucleus. Sturtevant also pointed out crossover interference. His summary is clear and succinct:

> It has been found possible to arrange six sex-linked factors in *Drosophila* in a linear series, using the number of crossovers per one hundred cases as an index of the distance between any two factors. This scheme gives consistent results, in the main.
>
> A source of error in predicting the strength of association between untried factors is found in double crossing over. The occurrence of this phenomenon is demonstrated, and it is shown not to occur as often as would be expected from a purely mathematical point of view, but the conditions governing its frequency are as yet not worked out.
>
> These results . . . form a new argument in favor of the chromosome view of inheritance, since they strongly indicate that the factors investigated are arranged in a linear series, at least mathematically.



Alfred H. Sturtevant (1891–1970). (Courtesy of the Archives, California Institute of Technology.)

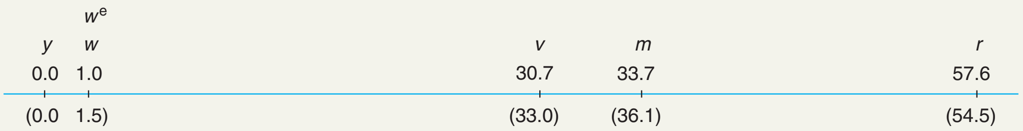


**Figure 1** The first chromosomal linkage map. Five loci in *Drosophila melanogaster* are mapped to the X chromosome. The numbers in parentheses are the more accurately mapped distances recognized today. We also show today's allelic designations rather than Sturtevant's original nomenclature. (Data from Sturtevant. "The linear arrangement of six sex-linked factors in *Drosophila,* as shown by their mode of association," *Journal of Experimental Zoology,* 14:43–59, 1913.)

Of twenty-eight offspring examined, all were consistent with the predictions of the Punnett square in figure 6.14. Those of class 8 (lower right box) with the colored, waxy phenotype all had a knobbed interchange chromosome as well as a normal homologue. Those with the colorless, waxy phenotype (class 4) had a knobless interchange chromosome. All of the colored, nonwaxy phenotypes (classes 5, 6, and 7) had a knobbed, normal chromosome, which indicated that only classes 5 and 6 were in the sample. Of the two that were tested, both were *WxWx,* indicating that they were of class 5. The remaining classes (1, 2, and 3) were of the colorless, nonwaxy phenotype. All were knobless. Of those that contained only normal chromosomes, some were

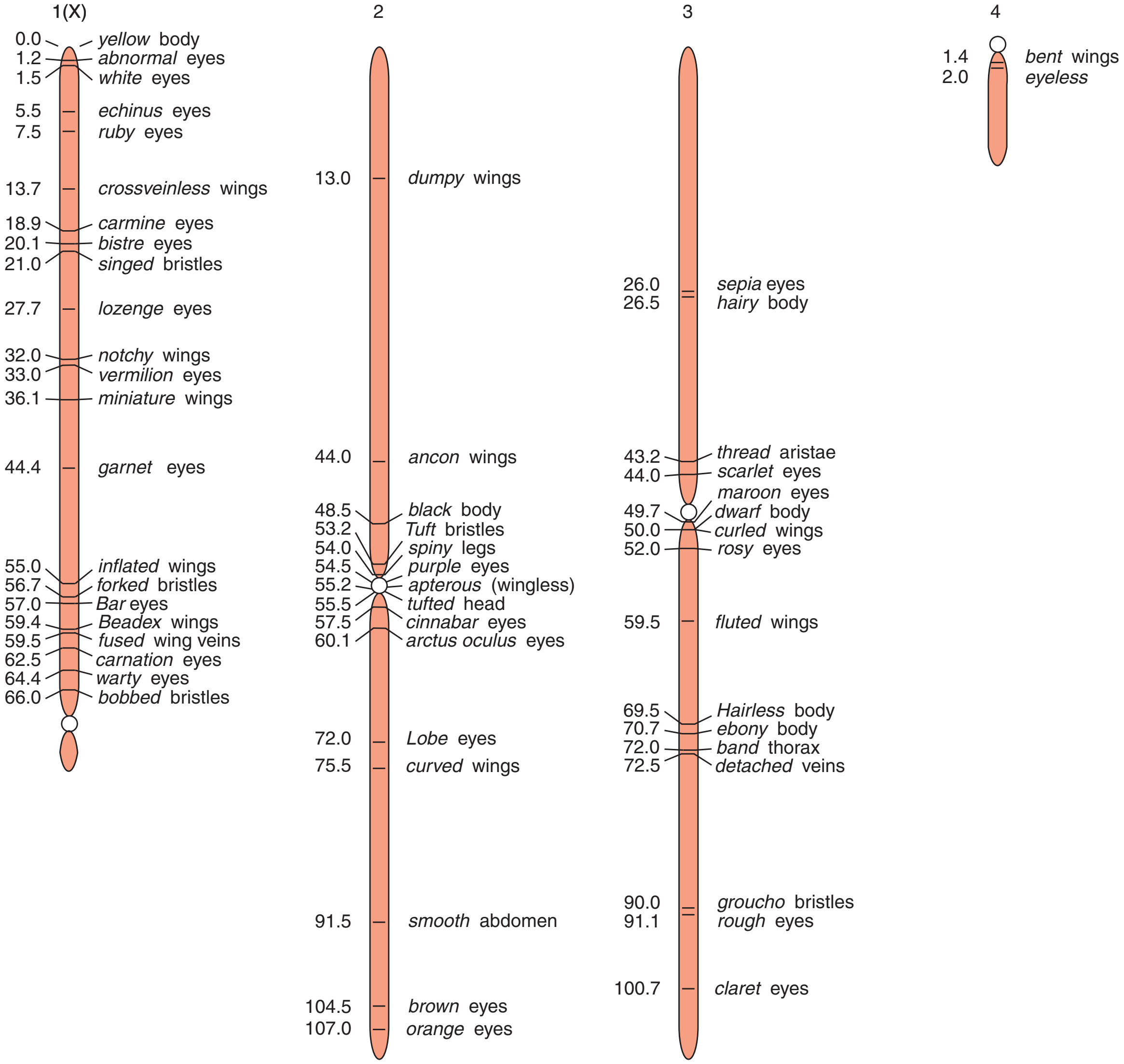

1(X)
0.0 yellow body
1.2 abnormal eyes
1.5 white eyes
5.5 echinus eyes
7.5 ruby eyes
13.7 crossveinless wings
18.9 carmine eyes
20.1 bistre eyes
21.0 singed bristles
27.7 lozenge eyes
32.0 notchy wings
33.0 vermilion eyes
36.1 miniature wings
44.4 garnet eyes
55.0 inflated wings
56.7 forked bristles
57.0 Bar eyes
59.4 Beadex wings
59.5 fused wing veins
62.5 carnation eyes
64.4 warty eyes
66.0 bobbed bristles
2
13.0 dumpy wings
44.0 ancon wings
48.5 black body
53.2 Tuft bristles
54.0 spiny legs
54.5 purple eyes
55.2 apterous (wingless)
55.5 tufted head
57.5 cinnabar eyes
60.1 arctus oculus eyes
72.0 Lobe eyes
75.5 curved wings
91.5 smooth abdomen
104.5 brown eyes
107.0 orange eyes
3
26.0 sepia eyes
26.5 hairy body
43.2 thread aristae
44.0 scarlet eyes
maroon eyes
49.7 dwarf body
50.0 curled wings
52.0 rosy eyes
59.5 fluted wings
69.5 Hairless body
70.7 ebony body
72.0 band thorax
72.5 detached veins
90.0 groucho bristles
91.1 rough eyes
100.7 claret eyes
4
1.4 bent wings
2.0 eyeless

**Figure 6.13** Partial map of the chromosomes of *Drosophila melanogaster.* The centromere is marked by an open circle.
(From C. Bridges, "Salavary Chromosome Maps," *Journal of Heredity,* 26:60–64, 1935. Reprinted with permission of Oxford University Press.)

*WxWx* (class 1) and some were heterozygotes (*Wxwx*, class 2). Of those containing interchange chromosomes, two were heterozygous, representing class 3. Two were homozygous, *WxWx*, yet interchange-normal heteromorphs. These represent a crossover in the region between the waxy locus and the extra piece of chromatin, producing a knobless-*c*-*Wx*-extra-piece chromosome. When combined with a *c*-*Wx*-normal chromosome, these would give these anomalous genotypes. The sample size was not large enough to pick up the reciprocal event. Creighton and McClintock concluded: "Pairing chromosomes, heteromorphic in two regions, have been shown to exchange parts at the same time they exchange genes assigned to these regions."

## HAPLOID MAPPING (TETRAD ANALYSIS)

For *Drosophila* and other diploid eukaryotes, the genetic analysis considered earlier in this chapter is referred to as **random strand analysis.** Sperm cells, each of which carry only one chromatid of a meiotic tetrad, unite with

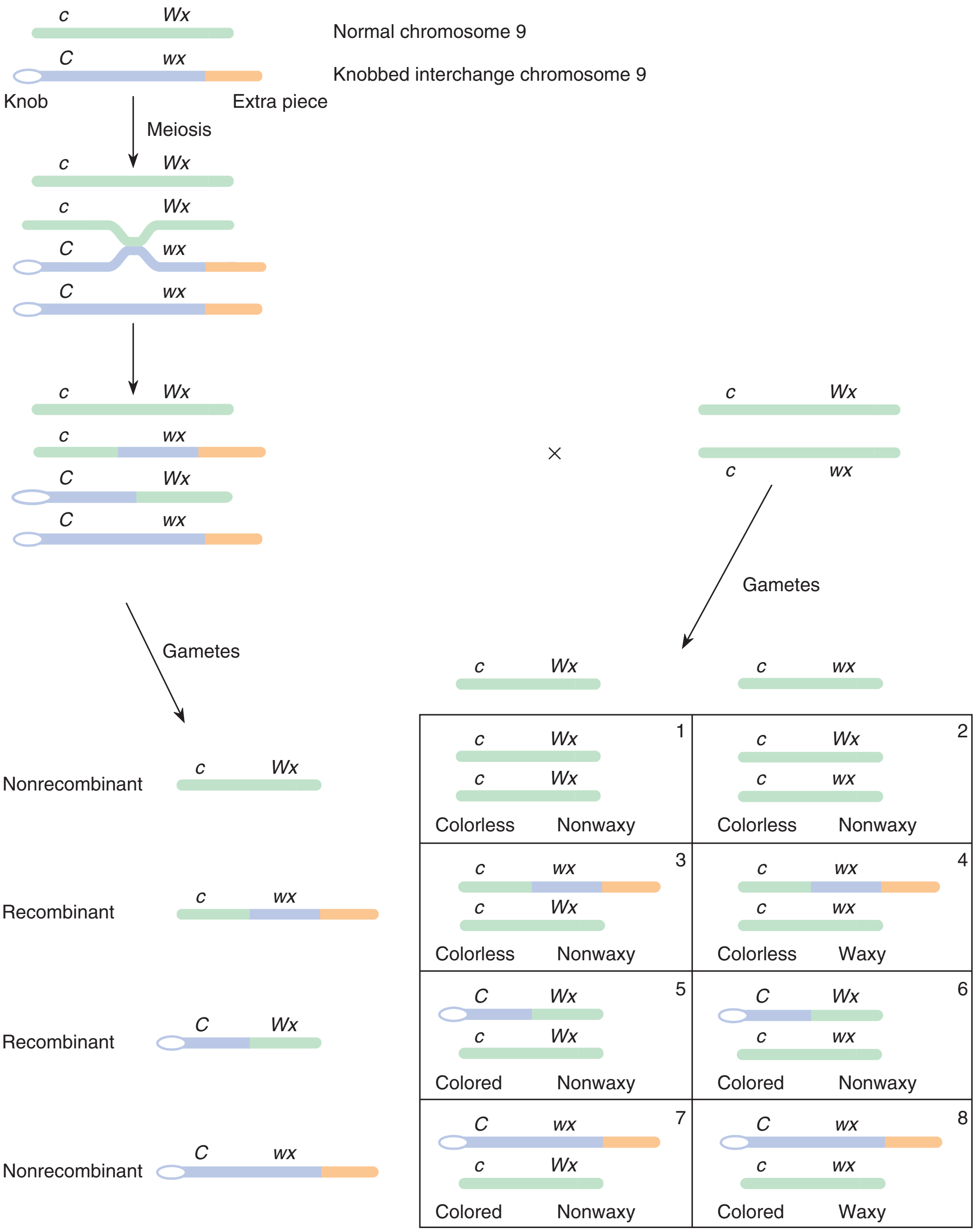

c Wx
Normal chromosome 9
C wx
Knobbed interchange chromosome 9
Knob
Extra piece
Meiosis
c Wx
c Wx
C wx
C wx
c Wx
c wx
C Wx
C wx
×
c Wx
c wx
Gametes
Gametes
c Wx
c wx
Nonrecombinant c Wx
Recombinant c wx
Recombinant C Wx
Nonrecombinant C wx
1 c Wx c Wx Colorless Nonwaxy
2 c Wx c wx Colorless Nonwaxy
3 c wx c Wx Colorless Nonwaxy
4 c wx c wx Colorless Waxy
5 C Wx c Wx Colored Nonwaxy
6 C Wx c wx Colored Nonwaxy
7 C wx c Wx Colored Nonwaxy
8 C wx c wx Colored Waxy

**Figure 6.14** Creighton and McClintock's experiment in maize demonstrated that genetic crossover correlates with cytological crossing over.

eggs, which also carry only one chromatid from a tetrad. Thus, zygotes are the result of the random uniting of chromatids.

Fungi of the class Ascomycetes retain the four haploid products of meiosis in a sac called an **ascus.** These organisms provide a unique opportunity to look at the total products of meiosis in a tetrad. Having the four products of meiosis allowed geneticists to determine such basics as the reciprocity of crossing over and the fact that DNA replication occurs before crossing over. Different techniques are used for these analyses. We will look at two fungi, the common baker's yeast, *Saccharomyces cerevisiae,* and pink bread mold, *Neurospora crassa,* both of which retain the products of meiosis as **ascospores.**

## *Phenotypes of Fungi*

At this point, you might wonder what phenotypes fungi such as yeast and *Neurospora* express. In general, microorganisms have phenotypes that fall into three broad categories: colony morphology, drug resistance, and nutritional requirements. Many microorganisms can be cultured in petri plates or test tubes that contain a supporting medium such as agar, to which various substances can be added (fig. 6.15). Wild-type *Neurospora,* the familiar pink bread mold, generally grows in a filamentous form, whereas yeast tends to form colonies. Various mutations exist that change colony morphology. In yeast, the *ade* gene causes the colonies to be red. In *Neurospora,* fluffy (*fl*), tuft (*tu*), dirty (*dir*), and colonial (*col4*) are all mutants of the basic growth form. In addition, wild-type *Neurospora* is sensitive to the sulfa drug sulfonamide, whereas one of its mutants (*Sfo*) actually requires sulfonamide in order to survive and grow. Yeast shows similar sensitivities to antifungal agents.

Nutritional-requirement phenotypes provide great insight not only into genetic analysis but also into the biochemical pathways of metabolism, as mentioned in chapter 2. Wild-type *Neurospora* can grow on a medium containing only sugar, a nitrogen source, some organic acids and salts, and the vitamin biotin. This is referred to as **minimal medium.** However, several different mutant types, or strains, of **Neurospora** cannot grow on this minimal medium until some essential nutrient is added. For example, one mutant strain will not grow on minimal medium, but will grow if one of the amino acids, arginine, is added (fig. 6.16). From this we can infer that the wild-type has a normal, functional enzyme in the synthetic pathway of arginine. The arginine-requiring mutant has an allele that specifies an enzyme that is incapable of converting one of the intermediates in the pathway directly into arginine or into one of the precursors to arginine. We can see that if the synthetic pathway is long, many different loci may have alleles that cause the strain to require arginine (fig. 6.17). This, in fact, happens, and the different loci are usually named $arg_1$, $arg_2$, and so on. There are numerous biosynthetic pathways in yeast and *Neurospora,* and mutants exhibit many different nutritional requirements. Mutants can be induced experimentally by radiation or by chemicals and other treatments. These, then, are the tools we use to analyze and map the chromosomes of microorganisms, including yeast and *Neurospora.* These techniques are expanded on in the next chapter.

## *Unordered Spores (Yeast)*

Baker's, or budding, yeast, *Saccharomyces cerevisiae,* exists in both a haploid and diploid form (fig. 6.18). The haploid form usually forms under nutritional stress (starvation). When better conditions return, haploid cells of the two sexes, called **a** and α **mating types,** fuse to form the diploid. (Mating types are generally the result of a one-locus, two-allele genetic system that determines that only opposite mating types can fuse. We discuss this system in more detail in chapter 16.) The haploid is again established by meiosis under starvation conditions. In yeast, all the products of meiosis are contained in the ascus. Let us look at a mapping problem, using the *a* and *b* loci for convenience.

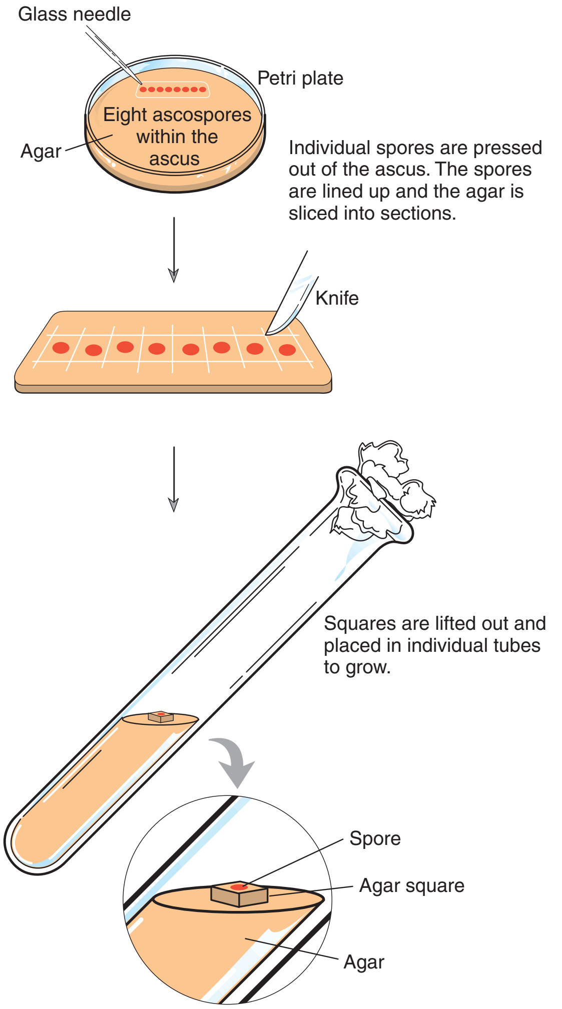

Glass needle
Petri plate
Eight ascospores within the ascus
Agar
Individual spores are pressed out of the ascus. The spores are lined up and the agar is sliced into sections.
Knife
Squares are lifted out and placed in individual tubes to grow.
Spore
Agar square
Agar

**Figure 6.15** Spore isolation technique in *Neurospora.*

When an *ab* spore (or gamete) fuses with an $a^+b^+$ spore (or gamete), and the diploid then undergoes meiosis, the spores can be isolated and grown as haploid colonies, which are then observed for the phenotypes the two loci control. Only three patterns can occur (table 6.4).

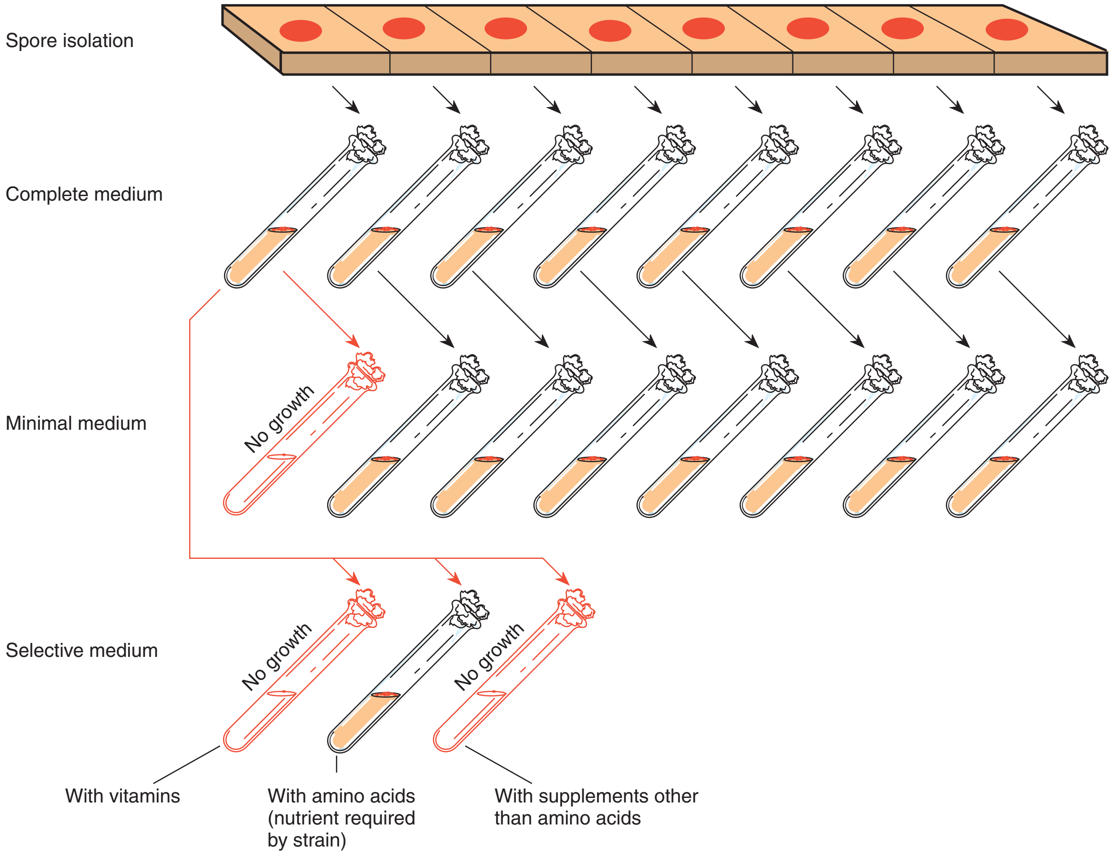


**Figure 6.16** Isolation of nutritional-requirement mutants in *Neurospora.*

Class 1 has two types of spores, which are identical to the parental haploid spores. This ascus type is, therefore, referred to as a **parental ditype (PD).** The second class also has only two spore types, but they are recombinants. This ascus type is referred to as a **nonparental ditype (NPD).** The third class has all four possible spore types and is referred to as a **tetratype (TT).**

All three ascus types can be generated whether or not the two loci are linked. As figure 6.19 shows, if the loci are linked, parental ditypes come from the lack of a crossover, whereas nonparental ditypes come about from four-strand double crossovers (double crossovers involving all four chromatids). We should thus expect parental ditypes to be more numerous than nonparental ditypes for linked loci. However, if the loci are not linked, both parental and nonparental ditypes come about through independent assortment—they should occur in equal frequencies. We can therefore determine whether the loci are linked by comparing parental ditypes and nonparental ditypes. In table 6.4, the parental ditypes greatly outnumber the nonparental ditypes; the two loci are, therefore, linked. What is the map distance between the loci?

A return to figure 6.19 shows that in a nonparental ditype, all four chromatids are recombinant, whereas in a tetratype, only half the chromatids are recombinant. Remembering that 1% recombinant offspring equals 1 map unit, we can use the following formula:

$$\text{map units} = \frac{(1/2)\text{ the number of TT asci} + \text{the number of NPD asci}}{\text{total number of asci}} \times 100$$

Thus, for the data of table 6.4,

$$\begin{matrix}\text{map}\\ \text{units}\end{matrix} = \frac{(1/2)20 + 5}{100} \times 100 = \frac{10 + 5}{100} \times 100 = 15$$

## *Ordered Spores (*Neurospora*)*

Unlike yeast, *Neurospora* has ordered spores; *Neurospora's* life cycle is shown in figure 6.20. Fertilization takes place within an immature fruiting body after a

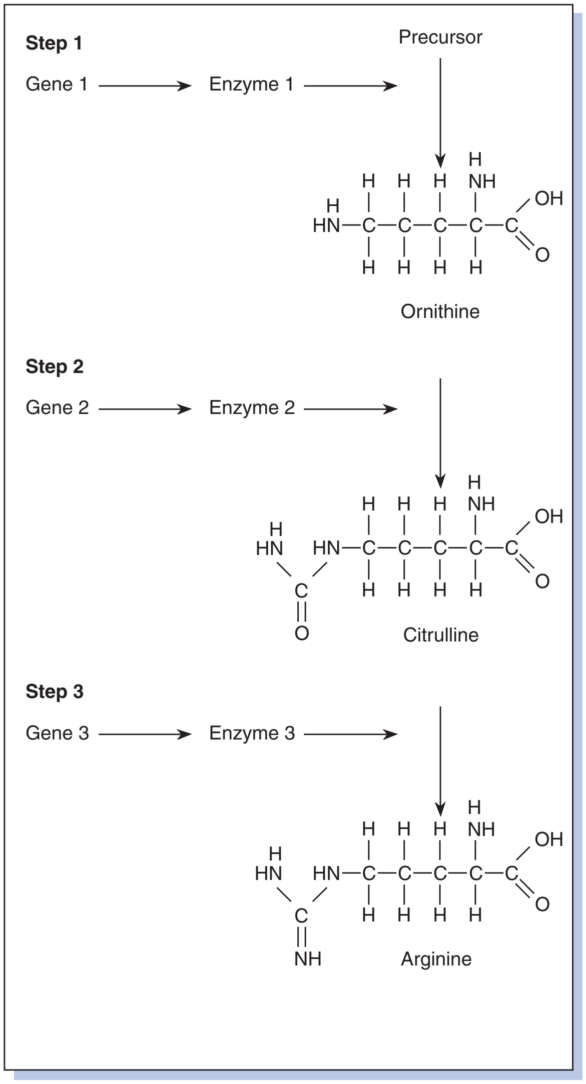

Step 1
Gene 1
Enzyme 1
Precursor
Ornithine
Step 2
Gene 2
Enzyme 2
Citrulline
Step 3
Gene 3
Enzyme 3
Arginine

**Figure 6.17** Arginine biosynthetic pathway of *Neurospora.*

**Table 6.4** **The Three Ascus Types in Yeast Resulting from Meiosis in a Dihybrid, $aa^{+}bb^{+}$**

| 1 (PD) | 2 (NPD) | 3 (TT) |
|---|---|---|
| $ab$ | $ab^{+}$ | $ab$ |
| $ab$ | $ab^{+}$ | $ab^{+}$ |
| $a^{+}b^{+}$ | $a^{+}b$ | $a^{+}b$ |
| $a^{+}b^{+}$ | $a^{+}b$ | $a^{+}b^{+}$ |
| 75 | 5 | 20 |

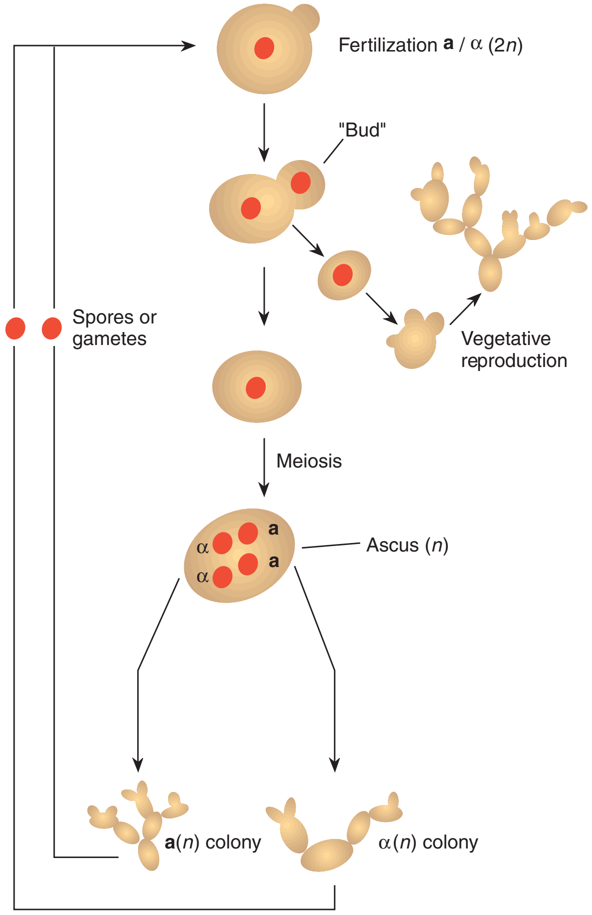
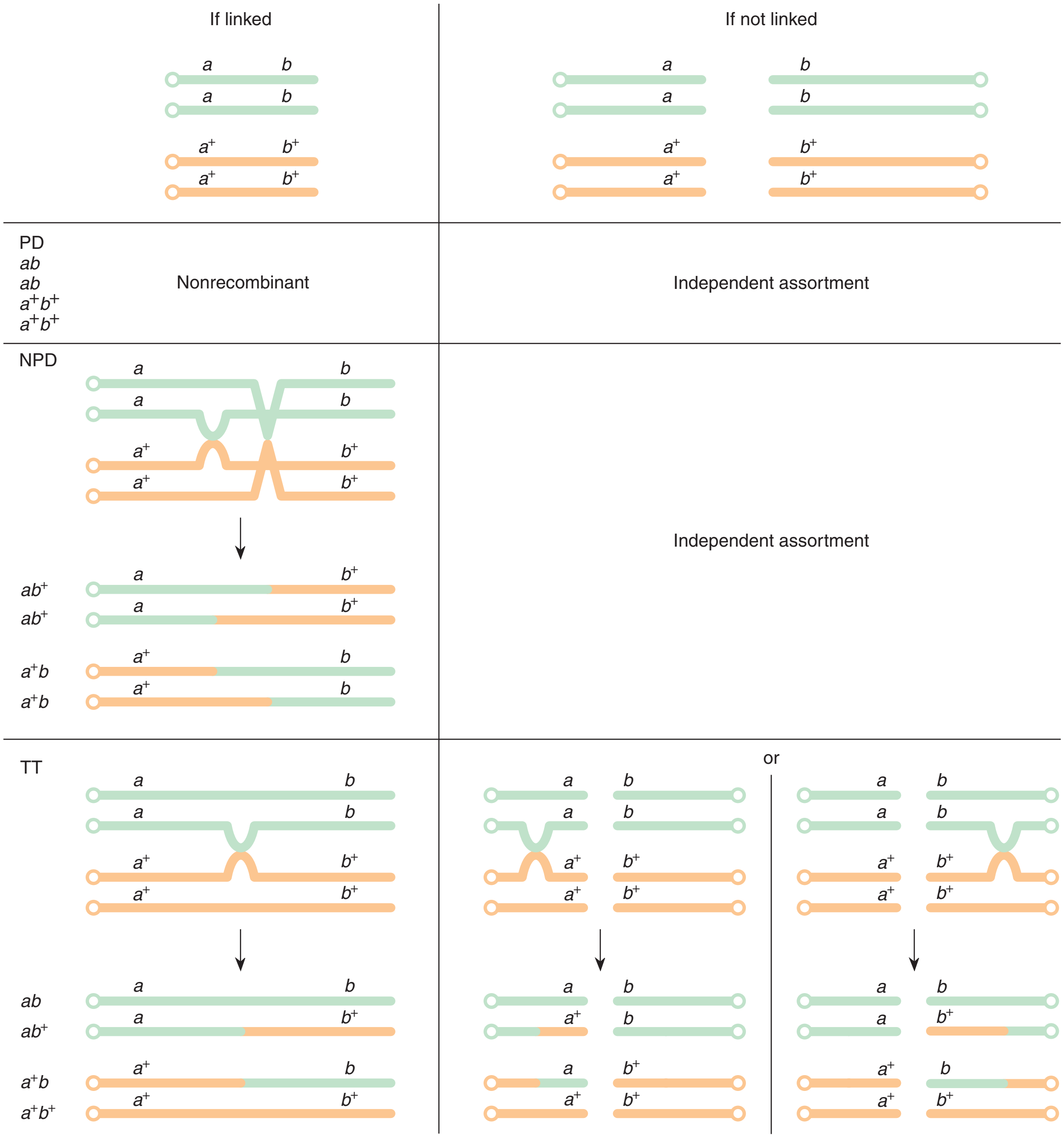

Fertilization **a** / α (2*n*)
"Bud"
Vegetative reproduction
Spores or gametes
Meiosis
α
α
a
a
Ascus (*n*)
**a**(*n*) colony
α(*n*) colony

**Figure 6.18** Life cycle of yeast. Mature cells are mating types **a** or α; *n* is the haploid stage; 2*n* is diploid.

spore or filament of one mating type contacts a special filament extending from the fruiting body of the opposite mating type (mating types are referred to as *A* and *a*). The zygote's nucleus undergoes meiosis without any intervening mitosis. Unlike yeast, *Neurospora* does not have a diploid phase in its life cycle. Rather, it undergoes meiosis immediately after the diploid nuclei form.

Since the *Neurospora* ascus is narrow, the meiotic spindle is forced to lie along the cell's long axis. The two nuclei then undergo the second meiotic division, which is also oriented along the long axis of the ascus. The result is that the spores are ordered according to their centromeres (fig. 6.21). That is, if we label one centromere *A* and the other *a,* for the two mating types, a tetrad at meiosis I will consist of one *A* and one *a* centromere. At the end of meiosis in *Neurospora,* the four ascospores are in the order *A A a a* or *a a A A* in regard to centromeres. (We talk more simply of centromeres rather than chromosomes or chromatids because of the complications that



If linked
If not linked
a
b
a
b
a+
b+
a+
b+
PD
ab
ab
a+b+
a+b+
Nonrecombinant
Independent assortment
NPD
ab+
ab+
a+b
a+b
Independent assortment
TT
or
ab
ab+
a+b
a+b+

**Figure 6.19** Formation of parental ditype (PD), nonparental ditype (NPD), and tetratype (TT) asci in a dihybrid yeast by linkage or independent assortment at meiosis. Open circles are centromeres.

crossing over adds. A type *A* centromere is always a type *A* centromere, whereas, due to crossing over, a chromosome attached to that centromere may be partly from the type *A* parent and partly from the type *a* parent.)

Before the ascospores mature in *Neurospora,* a mitosis takes place in each nucleus so that four pairs rather than just four spores are formed. In the absence of phenomena such as mutation or gene conversion, to be discussed later in the book, pairs are always identical (fig. 6.21). As we will see in a moment, because of the ordered spores, we can map loci in *Neurospora* in relation to their centromeres.

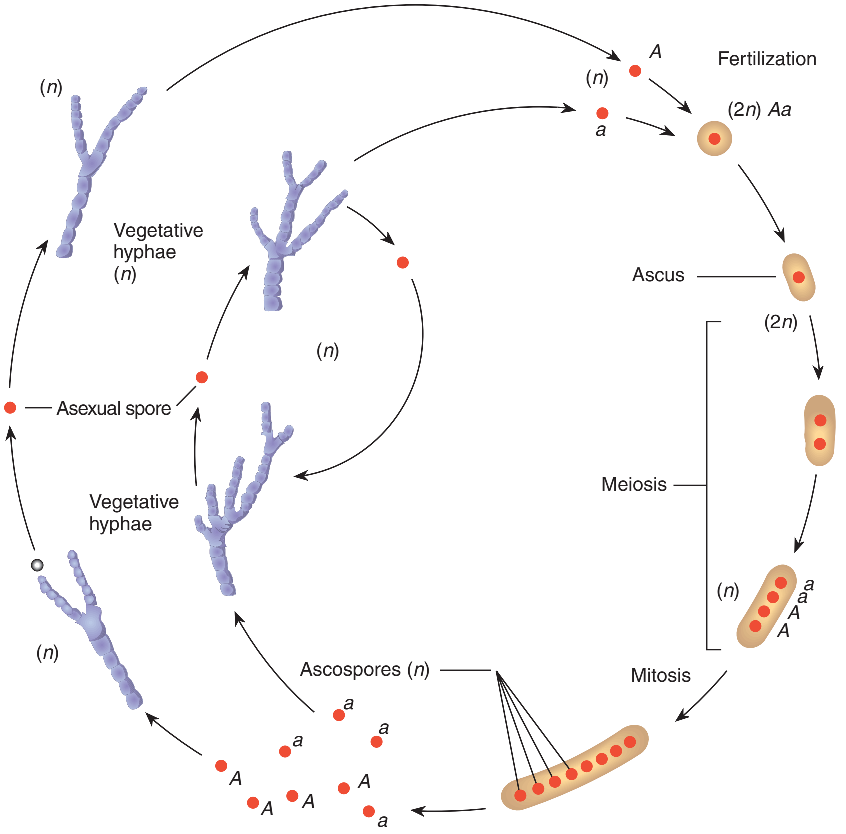


**Figure 6.20** Life cycle of *Neurospora. A* and *a* are mating types; *n* is a haploid stage; 2*n* is diploid.

### *First and Second Division Segregation*

Recall that there is a 4:4 segregation of the centromeres in the ascus of *Neurospora.* Two kinds of patterns appear among the loci on these chromosomes. These patterns depend on whether there was a crossover between the locus and its centromere (fig. 6.22). If there was no crossover between the locus and its centromere, the allelic pattern is the same as the centromeric pattern, which is referred to as **first-division segregation (FDS),** because the alleles separate from each other at meiosis I. If, however, a crossover has occurred between the locus and its centromere, patterns of a different type emerge (2:4:2 or 2:2:2:2), each of which is referred to as **second-division segregation (SDS).** Because the spores are ordered, the centromeres always follow a first-division segregation pattern. Hence, we should be able to map the distance of a locus to its centromere. Under the simplest circumstances (fig. 6.22), every second-division segregation configuration has four recombinant and four nonrecombinant chromatids (spores). Thus, half of the chromatids (spores) in a second-division segregation ascus are recombinant. Therefore, since 1% recombinant chromatids equal 1 map unit,

$$\text{map distance} = \frac{(1/2)\text{ the number of SDS asci}}{\text{total number of asci}} \times 100$$

An example using this calculation appears in table 6.5.

Three-point crosses in *Neurospora* can also be examined. Let us map two loci and their centromere. For simplicity, we will use the *a* and *b* loci. Dihybrids are formed from fused mycelia ($ab \times a^+b^+$), which then undergo meiosis. One thousand asci are analyzed, keeping the spore order intact. Before presenting the data, we should consider how to group them. Since each locus can show six different patterns (fig. 6.22), two loci scored together should give thirty-six possible spore arrangements (6 × 6). Some thought, however, tells us that many of these patterns are really random variants of each other. The tetrad in meiosis is a three-dimensional entity rather than a flat, four-rod object, as it is usually drawn. At the first meiotic division, either centromere can go to the left *or* the right, and when centromeres split at the second mei-

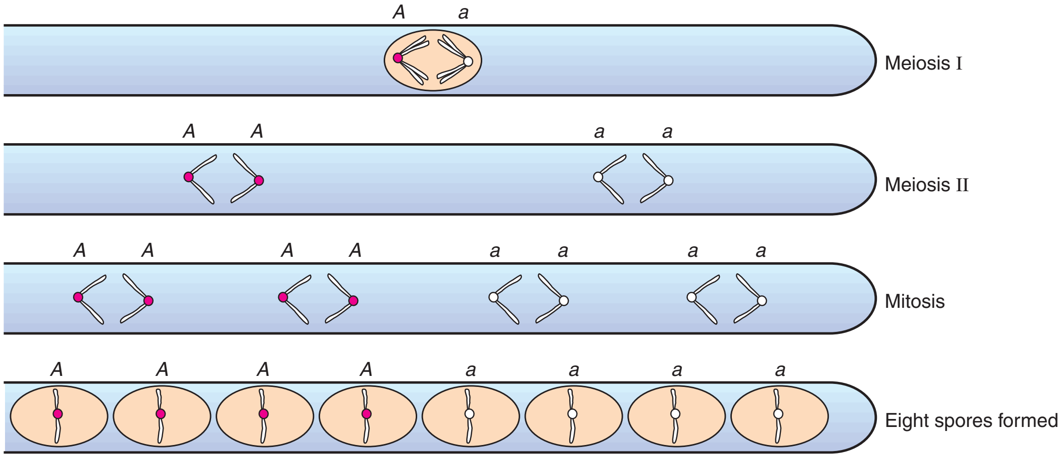


**Figure 6.21** Meiosis in *Neurospora.* Although *Neurospora* has seven pairs of chromosomes at meiosis, only one pair is shown. *A* and *a,* the two mating types, represent the two centromeres of the tetrad.

otic division, movement within the future half-ascus (the four spores to the left or the four spores to the right) is also random. Thus, one genetic event can produce up to eight "different" patterns. For example, consider the arrangements figure 6.23 shows, in which a crossover occurs between the *a* and *b* loci. All eight arrangements, producing the ascus patterns of table 6.6, are equally likely. The thirty-six possible patterns then reduce to only the seven unique patterns shown in table 6.7. Note also that these asci can be grouped into the three types of asci found in yeast with unordered spores: parental ditypes, nonparental ditypes, and tetratypes. Had we not had the order of the spores from the asci, that would, in fact, be the only way we could score the asci (see the bottom of table 6.7).

### Gene Order

We can now determine the distance from each locus to its centromere and the linkage arrangement of the loci if they are both linked to the same centromere. We can establish by inspection that the two loci are linked to each

**Table 6.5** **Genetic Patterns Following Meiosis in an $a^+a$ Heterozygous *Neurospora* (Ten Asci Examined)**

| Spore Number | Ascus Number | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | $a$ | $a$ | $a^+$ | $a$ | $a$ | $a^+$ | $a$ | $a^+$ | $a^+$ | $a^+$ |
| 2 | $a$ | $a$ | $a^+$ | $a$ | $a$ | $a^+$ | $a$ | $a^+$ | $a^+$ | $a^+$ |
| 3 | $a$ | $a$ | $a^+$ | $a^+$ | $a^+$ | $a^+$ | $a$ | $a$ | $a$ | $a^+$ |
| 4 | $a$ | $a$ | $a^+$ | $a^+$ | $a^+$ | $a^+$ | $a$ | $a$ | $a$ | $a^+$ |
| 5 | $a^+$ | $a^+$ | $a$ | $a^+$ | $a$ | $a$ | $a^+$ | $a$ | $a^+$ | $a$ |
| 6 | $a^+$ | $a^+$ | $a$ | $a^+$ | $a$ | $a$ | $a^+$ | $a$ | $a^+$ | $a$ |
| 7 | $a^+$ | $a^+$ | $a$ | $a$ | $a^+$ | $a$ | $a^+$ | $a^+$ | $a$ | $a$ |
| 8 | $a^+$ | $a^+$ | $a$ | $a$ | $a^+$ | $a$ | $a^+$ | $a^+$ | $a$ | $a$ |
| | FDS | FDS | FDS | SDS | SDS | FDS | FDS | SDS | SDS | FDS |

*Note:* Map distance (*a* locus to centromere) = (1/2)% SDS
= (1/2) 40%
= 20 map units

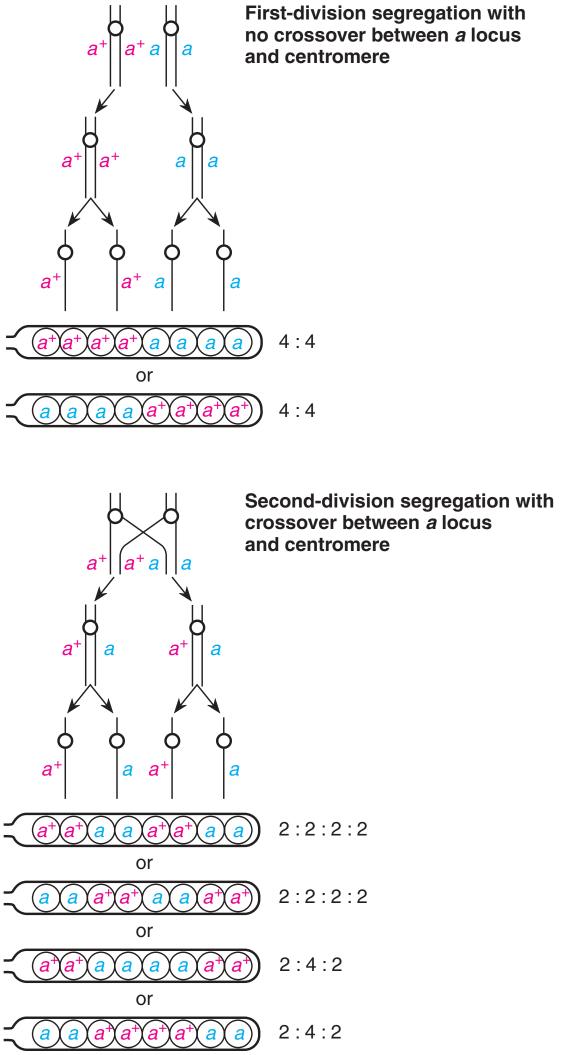

First-division segregation with no crossover between *a* locus and centromere
$a^+$
$a^+$
a
a
4 : 4
or
4 : 4
Second-division segregation with crossover between *a* locus and centromere
2 : 2 : 2 : 2
or
2 : 2 : 2 : 2
or
2 : 4 : 2
or
2 : 4 : 2

**Figure 6.22** The six possible *Neurospora* ascospore patterns in respect to one locus.

**Table 6.6** Eight of the Thirty-Six Possible Spore Patterns in *Neurospora* Scored for Two Loci, *a* and *b* (All Random Variants of the Same Genetic Event)

| Spore Number | Ascus Number 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | $ab$ | $ab^+$ | $ab$ | $ab^+$ | $a^+b^+$ | $a^+b^+$ | $a^+b$ | $a^+b$ |
| 2 | $ab$ | $ab^+$ | $ab$ | $ab^+$ | $a^+b^+$ | $a^+b^+$ | $a^+b$ | $a^+b$ |
| 3 | $ab^+$ | $ab$ | $ab^+$ | $ab$ | $a^+b$ | $a^+b$ | $a^+b^+$ | $a^+b^+$ |
| 4 | $ab^+$ | $ab$ | $ab^+$ | $ab$ | $a^+b$ | $a^+b$ | $a^+b^+$ | $a^+b^+$ |
| 5 | $a^+b$ | $a^+b$ | $a^+b^+$ | $a^+b^+$ | $ab^+$ | $ab$ | $ab^+$ | $ab$ |
| 6 | $a^+b$ | $a^+b$ | $a^+b^+$ | $a^+b^+$ | $ab^+$ | $ab$ | $ab^+$ | $ab$ |
| 7 | $a^+b^+$ | $a^+b^+$ | $a^+b$ | $a^+b$ | $ab$ | $ab^+$ | $ab$ | $ab^+$ |
| 8 | $a^+b^+$ | $a^+b^+$ | $a^+b$ | $a^+b$ | $ab$ | $ab^+$ | $ab$ | $ab^+$ |

**Table 6.7** The Seven Unique Classes of Asci Resulting from Meiosis in a Dihybrid *Neurospora*, $ab/a^+b^+$

| Spore Number | Ascus Number 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | $ab$ | $ab^+$ | $ab$ | $ab$ | $ab$ | $ab^+$ | $ab$ |
| 2 | $ab$ | $ab^+$ | $ab$ | $ab$ | $ab$ | $ab^+$ | $ab$ |
| 3 | $ab$ | $ab^+$ | $ab^+$ | $a^+b$ | $a^+b^+$ | $a^+b$ | $a^+b^+$ |
| 4 | $ab$ | $ab^+$ | $ab^+$ | $a^+b$ | $a^+b^+$ | $a^+b$ | $a^+b^+$ |
| 5 | $a^+b^+$ | $a^+b$ | $a^+b^+$ | $a^+b^+$ | $a^+b^+$ | $a^+b$ | $a^+b$ |
| 6 | $a^+b^+$ | $a^+b$ | $a^+b^+$ | $a^+b^+$ | $a^+b^+$ | $a^+b$ | $a^+b$ |
| 7 | $a^+b^+$ | $a^+b$ | $a^+b$ | $ab^+$ | $ab$ | $ab^+$ | $ab^+$ |
| 8 | $a^+b^+$ | $a^+b$ | $a^+b$ | $ab^+$ | $ab$ | $ab^+$ | $ab^+$ |
|  | 729 | 2 | 101 | 9 | 150 | 1 | 8 |
| SDS for *a* locus |  |  |  | 9 | 150 | 1 | 8 |
| SDS for *b* locus |  |  | 101 |  | 150 | 1 | 8 |
| Unordered: | PD | NPD | TT | TT | PD | NPD | TT |

other—and therefore to the same centromere—by examining classes 1 (parental ditype) and 2 (nonparental ditype) in table 6.7. If the two loci are unlinked, these two categories would represent two equally likely alternative events when no crossover takes place. Since category 1 represents almost 75% of all the asci, we can be sure the two loci are linked.

To determine the distance of each locus to the centromere, we calculate one-half the percentage of second-division segregation patterns for each locus. For the *a* locus, classes 4, 5, 6, and 7 are second-division segregation patterns. For the *b* locus, classes 3, 5, 6, and 7 are second-division segregation patterns. Therefore, the distances to the centromere, in map units, for each locus are

$$\text{for locus } a\text{: } (1/2)\,\frac{9 + 150 + 1 + 8}{1{,}000} \times 100$$

$$= 8.4 \text{ centimorgans}$$

$$\text{for locus } b\text{: } (1/2)\,\frac{101 + 150 + 1 + 8}{1{,}000} \times 100$$

$$= 13.0 \text{ centimorgans}$$

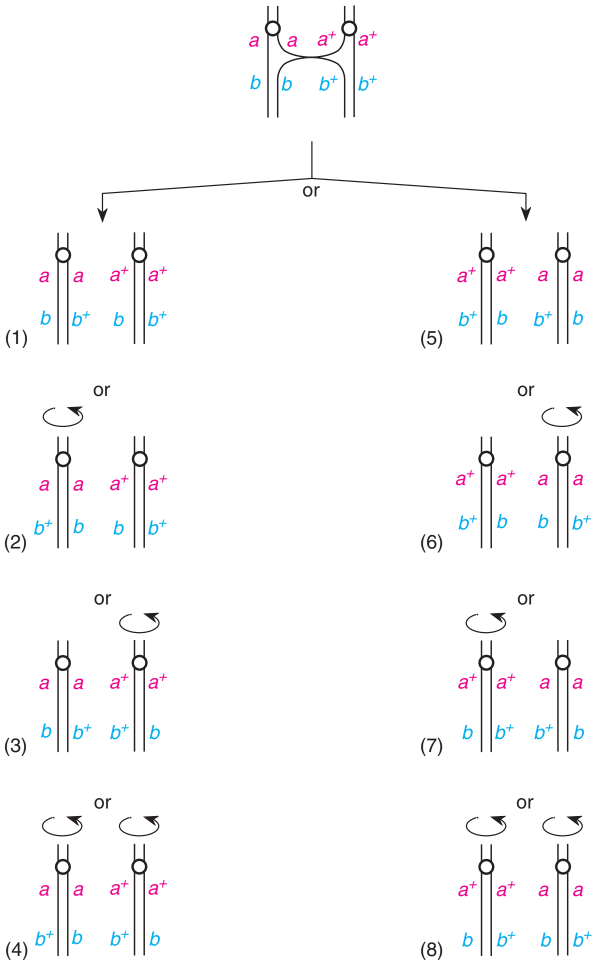


**Figure 6.23** The eight random arrangements possible when a single crossover occurs between the *a* and *b* loci in *Neurospora* (see table 6.6). *Circular arrows* represent the rotation of a centromere from its position in the original configuration.

It should now be possible to describe exactly what type of crossover event produced each of the seven classes in table 6.7.

Unfortunately, these two distances do not provide a unique determination of gene order. In figure 6.24, we see that two alternatives are possible: one has a map distance between the loci of 21.4 map units; the other has 4.6 map units between loci. How do we determine which of these is correct? The simplest way is to calculate the *a–b* distance using the unordered spore information. That is, the map distance is

$$\text{map units} = \frac{\text{(1/2) the number of TT asci} + \text{the number of NPD asci}}{\text{total number of asci}} \times 100$$

$$= \frac{(1/2)118 + 3}{1{,}000} \times 100 = 6.2$$

Since 6.2 map units is much closer to the *a–b* distance expected if both loci are on the same side of the centromere, we accept alternative 2 in figure 6.24.

A second way to choose between the alternatives in figure 6.24 is to find out what happens to the *b* locus when a crossover occurs between the *a* locus and its centromere. If the order in alternative 1 is correct, crossovers between the *a* locus and its centromere should have no effect on the *b* locus; if 2 is correct, most of the crossovers that move the *a* locus in relation to its centromere should also move the *b* locus.

Asci classes 4, 5, 6, and 7 include all the SDS patterns for the *a* locus. Of 168 asci, 150 (class 5) have similar SDS patterns for the *b* locus. Thus, 89% of the time, a crossover between the *a* locus and its centromere is also a crossover between the *b* locus and its centromere—compelling evidence in favor of alternative 2. (What form would the data take if alternative 1 were correct?)

In summary, mapping by tetrad analysis proceeds as follows. For both ordered and unordered spores, linkage is indicated by an excess of parental ditypes over nonparental ditypes. For unordered spores (yeast), the distance between two loci is one-half the number of

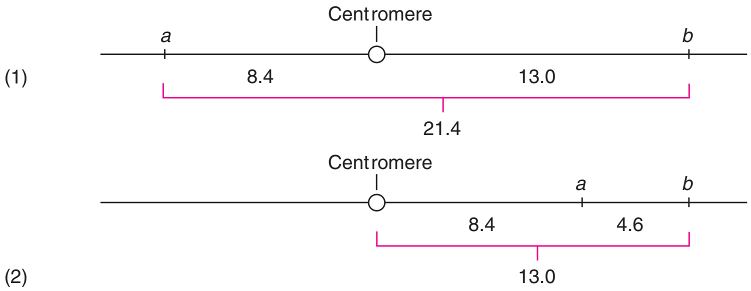


**Figure 6.24** Two possible arrangements of the *a* and *b* loci and their centromere. Distances are in map units.

tetratypes plus the number of nonparental ditypes, all divided by the total number of asci, expressed as a percentage. For ordered spores (*Neurospora*), the distance from a locus to its centromere is one-half the percentage of second-division segregants. Mapping the distance between two loci is similar to the process in unordered spores.

## SOMATIC (MITOTIC) CROSSING OVER

Crossing over is known to occur in somatic cells as well as during meiosis. It apparently occurs when two homologous chromatids come to lie next to each other and breakage and reunion follow, most likely as a consequence of DNA repair (see chapter 12). Unlike in meiosis, no synaptonemal complex forms. The occurrence of mitotic crossing over is relatively rare. In the fungus *Aspergillus nidulans,* mitotic crossing over occurs about once in every one hundred cell divisions.

Mitotic recombination was discovered in 1936 by Curt Stern, who noticed the occurrence of *twin spots* in fruit flies that were dihybrid for the yellow allele for body color (*y*) and the singed allele (*sn*) for bristle morphology (fig. 6.25). A twin spot could be explained by mitotic crossing over between the *sn* locus and its centromere (fig. 6.26). A crossover in the *sn*–*y* region would produce only a yellow spot, whereas a double crossover, one between *y* and *sn* and the other between *sn* and the centromere, would produce only a singed spot. (Verify this for yourself.) These three phenotypes were found in the relative frequencies expected. That is, given that the gene locations are drawn to scale in figure 6.26, we would expect double spots to be most common, followed by yellow spots, with singed spots rarest of all because they require a double crossover. This in fact occurred, and no other obvious explanation was consistent with these facts. Mitotic crossing over has been used in fungal genetics as a supplemental, or even a primary, method for determining linkage relations. Although gene orders are consistent between mitotic and meiotic mapping, relative distances are usually not, which is not totally unexpected. We know that neither meiotic nor mitotic crossing over is uniform along a chromosome. Apparently, the factors that cause deviation from uniformity differ in the two processes.

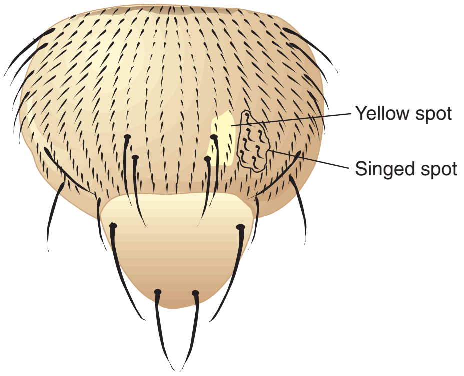
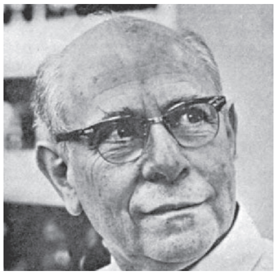


**Figure 6.25** Yellow and singed twin spots on the thorax of a female *Drosophila.*



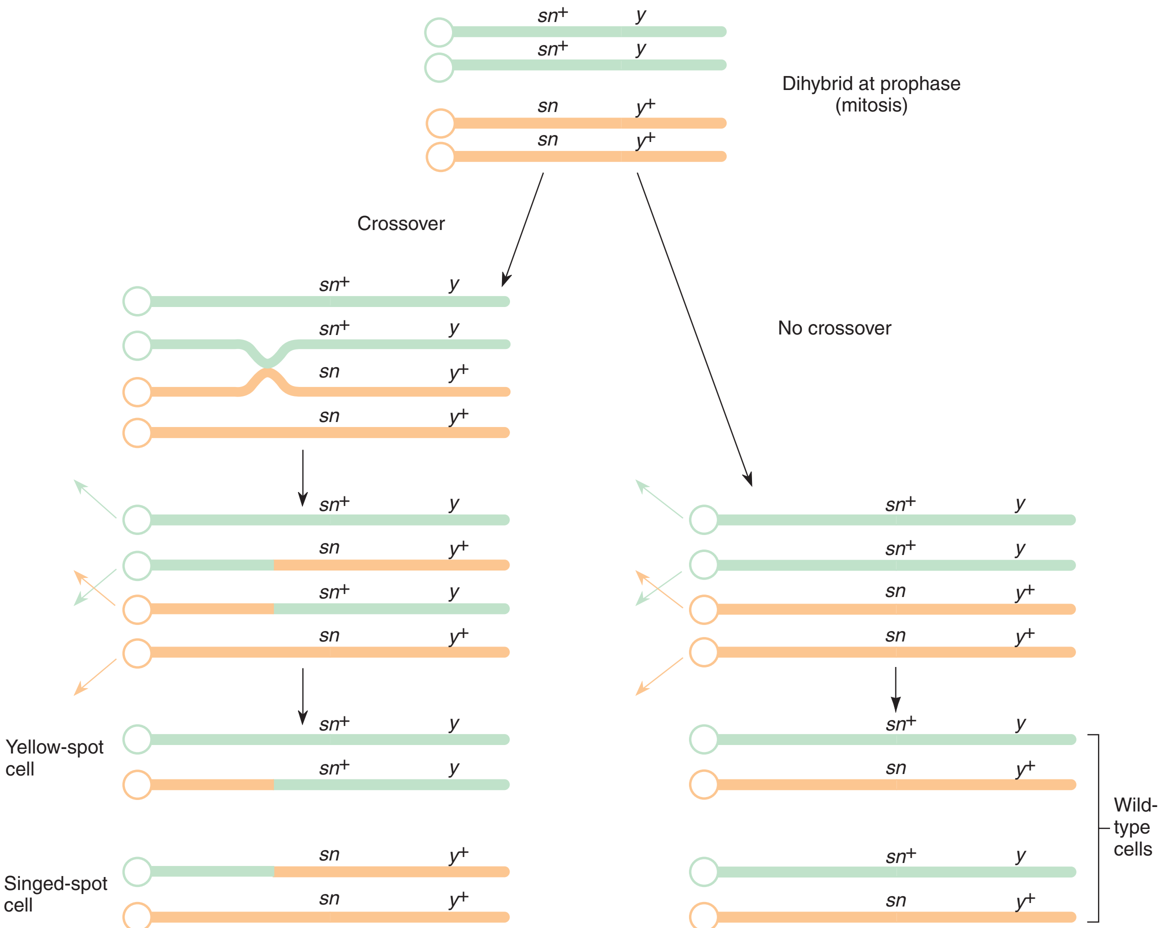
Curt Stern (1902–1981)
(Courtesy of the Science Council of Japan.)

## HUMAN CHROMOSOMAL MAPS

In theory, we can map human chromosomes as we would those of any other organism. Realistically, the problems mentioned earlier (the inability to make specific crosses coupled with the relatively small number of offspring) make these techniques of human chromosome mapping very difficult. However, some progress has been made based on pedigrees, especially in assigning genes to the X chromosome. As the pedigree analysis in the previous chapter has shown, X chromosomal traits have unique patterns of inheritance, and loci on the X chromosome are easy to identify. Currently over four hundred loci are known to be on the X chromosome. It has been estimated, by several different methods, that between fifty and one hundred thousand loci exist on human chromosomes. In later chapters, we will discuss several additional methods of human chromosomal mapping that use molecular genetic techniques.

### X Linkage

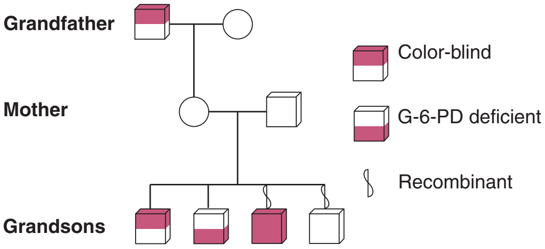
After determining that a human gene is X linked, the next problem is to determine the position of the locus on the X chromosome and the map units between loci. Sometimes we can do this with the proper pedigrees, if crossing over can be ascertained. An example of this “grand-



sn+ y
sn+ y
sn y+
sn y+
Dihybrid at prophase (mitosis)
Crossover
No crossover
Yellow-spot cell
Singed-spot cell
Wild-type cells

**Figure 6.26** Formation of twin spots by somatic crossing over.

father method" appears in figure 6.27. In this example, a grandfather has one of the traits in question (here, color blindness). We then find that he has a grandson who is glucose-6-phosphate dehydrogenase (G-6-PD) deficient. From this we can infer that the mother (of the grandson) was dihybrid for the two alleles in the *trans* configuration. That is, she received her color-blindness allele on one of her X chromosomes from her father, and she must have received the G-6-PD-deficiency allele on the other X chromosome from her mother (why?). Thus, the two sons on the left in figure 6.27 are nonrecombinant, and the two on the right are recombinant. Theoretically, we can determine map distance by simply totaling the recombinant grandsons and dividing by the total number of grandsons. Of course, the methodology would be the same if the grandfather were both color-blind and G-6-PD deficient. The mother would then be dihybrid in the *cis* configuration, and the sons would be tabulated in the reverse manner. The point is that the grandfather's phenotype gives us information that allows us to infer that the mother was dihybrid, as well as telling us the *cis-trans* arrangement of her alleles. We can then score her sons as either recombinant or nonrecombinant.



Grandfather
Mother
Grandsons
Color-blind
G-6-PD deficient
Recombinant

**Figure 6.27** "Grandfather method" of determining crossing over between loci on the human X chromosome. G-6-PD is glucose-6-phosphate dehydrogenase.

## Autosomal Linkage

From this we can see that it is relatively easy to map the X chromosome. The autosomes are another story. Since there are twenty-two autosomal linkage groups (twenty-two pairs of nonsex chromosomes), it is virtually impossible to determine from simple pedigrees which chromosome two loci are located on. Pedigrees can tell us if two loci are linked to each other, but not on which chromosome. In figure 6.28, the nail-patella syndrome includes, among other things, abnormal nail growth coupled with the absence or underdevelopment of kneecaps. It is a dominant trait. The male in generation II is dihybrid, with the *A* allele of the ABO blood type system associated with the nail-patella allele (*NPS1*) and the *B* allele with the normal nail-patella allele (*nps1*). Thus only one child in eight (III-5) is recombinant. Actually, the map distance is about 10%. In general, map distances appear greater in females than in males because more crossing over occurs in females (box 6.3).

We now turn our attention to the localization of loci to particular human chromosomes. The first locus that was definitely established to be on a particular autosome was the Duffy blood group on chromosome 1. This was ascertained in 1968 from a family that had a morphologically odd, or "uncoiled," chromosome 1. Inheritance in the Duffy blood group system followed the pattern of inheritance of the "uncoiled" chromosome. Real strides have been made since then. Two techniques, chromosomal banding and somatic-cell hybridization, have been crucial to autosomal mapping.

### Chromosomal Banding

Techniques were developed around 1970 that make use of certain histochemical stains that produce repeatable banding patterns on the chromosomes. For example, Giemsa staining is one such technique; the resulting bands are called **G-bands.** More detail on these techniques is presented in chapter 15. Before these techniques, human and other mammalian chromosomes were grouped into general size categories because of the difficulty of differentiating many of them. With banding techniques came the ability to identify each human chromosome in a karyotype (see fig. 5.1).

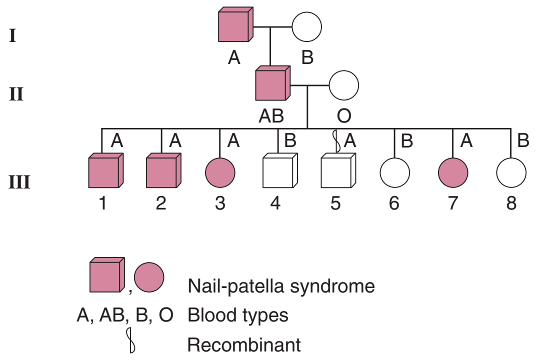


**Figure 6.28** Linkage of the nail-patella syndrome and ABO loci.

### Somatic-Cell Hybridization

The ability to distinguish each human chromosome is required to perform somatic-cell hybridization, in which human and mouse (or hamster) cells are fused in culture to form a hybrid. The fusion is usually mediated chemically with polyethylene glycol, which affects cell membranes; or with an inactivated virus, for example the Sendai virus, that is able to fuse to more than one cell at the same time. (The virus is able to do this because it has a lipid membrane derived from its host cells that easily fuses with new host cells. Because of this property, the virus can fuse to two cells close together, forming a cytoplasmic bridge between them that facilitates their fusion.) When two cells fuse, their nuclei are at first separate, forming a **heterokaryon,** a cell with nuclei from different sources. When the nuclei fuse, a hybrid cell is formed, and this hybrid tends to lose human chromosomes preferentially through succeeding generations. Upon stabilization, the result is a cell with one or more human chromosomes in addition to the original mouse or hamster chromosomal complement. Banding techniques allow the observer to recognize the human chromosomes. A geneticist looks for specific human phenotypes, such as enzyme products, and can then assign the phenotype to one of the human chromosomes in the cell line.

When cells are mixed together for hybridization, some cells do not hybridize. It is thus necessary to be able to select for study just those cells that are hybrids. One technique, originally devised by J. W. Littlefield in 1964, makes use of genetic differences in the way the cell lines synthesize DNA. Normally, in mammalian cells, aminopterin acts as an inhibitor of enzymes involved in DNA metabolism. Two enzymes, hypoxanthine phosphoribosyl transferase (HPRT) and thymidine kinase (TK), can bypass aminopterin inhibition by making use of secondary, or salvage, pathways in the cell. If hypoxanthine is provided, HPRT converts it to a purine, and if thymidine is provided, TK converts it to the nucleotide thymidylate. (Purines are converted to nucleotides and nucleotides are the subunits of DNA—see chapter 9.) Thus, normal cells in the absence of aminopterin synthesize DNA even if they lack HPRT activity ($\text{HPRT}^-$) or TK activity ($\text{TK}^-$). In the presence of aminopterin, $\text{HPRT}^-$ or $\text{TK}^-$ cells die. However, in the presence of aminopterin, $\text{HPRT}^+$ $\text{TK}^+$ cells can synthesize DNA and survive. Using this information, the following selection system was developed.

Mouse cells that have the phenotype of $\text{HPRT}^+$ $\text{TK}^-$ are mixed with human cells that have the phenotype of

## BOX 6.3

### Experimental Methods

### Lod *Scores*

Human population geneticists can increase the accuracy of their linkage analysis by using a probability technique, developed by Newton Morton, called the ***lod* score method** (*L*og *Od*ds). The geneticist asks what the probability is of getting a particular pedigree assuming a particular recombination frequency (Θ), as compared with getting the same pedigree assuming independent assortment (Θ = 0.50). In other words, he or she calculates the ratio of the probability of genotypes in a family given a certain crossover frequency compared with the probability of those genotypes if the loci are unlinked. Logarithms are used for ease of calculation, and the parameter is called *z*, the *lod* score. Using this method, a researcher can try different crossover frequencies until the one giving the highest *lod* score is found.

For example, take the pedigree in figure 6.28. The father in generation II can have one of two allelic arrangements: *A NPS1/B nps1* or *A/B NPS1/nps1.* The former assumes linkage, whereas the latter does not. Our initial estimate of recombination, assuming linkage, was (1/8) × 100, or 12.5 map units. We now need to calculate the ratio of two probabilities:

$$z = \log \frac{\text{probability of birth sequence assuming 12.5 map units}}{\text{probability of birth sequence assuming independent assortment}}$$

Assuming 12.5 map units (or a probability of 0.125 of a crossover; Θ = 0.125), the probability of child III-1 is 0.4375. This child would be a nonrecombinant, so his probability of having the nail-patella syndrome and type A blood is half the probability of no crossover during meiosis, or (1 − 0.125)/2. We divide by two because there are two nonrecombinant types. This is the same probability for all children except III-5, whose probability of occurrence is 0.125/2 = 0.0625, since he is a recombinant. Thus, the numerator of the previous equation is $(0.4375)^7(0.0625)$.

If the two loci are not linked, then any genotype has a probability of 1/4, or 0.25. Thus, the sequence of the eight children has the probability of $(0.25)^8$. This is the denominator of the equation. Thus,

$$z = \log \frac{(0.4375)^7(0.0625)}{(0.25)^8}$$

$$z = \log [12.566] = 1.099$$

Any *lod* score greater than zero favors linkage. A *lod* score less than zero suggests that Θ has been underestimated. A *lod* of 3.0 or greater ($10^3$ or one thousand times more likely than independent assortment) is considered a strong likelihood of linkage. Thus, in our example, we have an indication of linkage with a recombination frequency of 0.125. Now we can calculate *lod* scores assuming other values of recombination, as table 1 does. You can see that the recombination frequency as calculated, 0.125 (12.5 map units), gives the highest *lod* score.

Newton E. Morton (1929– ).

(Courtesy of Dr. Newton E. Morton.)

**Table 1** ***Lod* Scores for the Cross in Figure 6.28**

| Recombination Frequency (Θ) | *Lod* Score |
|---|---|
| 0.05 | 0.951 |
| 0.10 | 1.088 |
| 0.125 | 1.099 |
| 0.15 | 1.090 |
| 0.20 | 1.031 |
| 0.25 | 0.932 |
| 0.30 | 0.801 |
| 0.35 | 0.643 |
| 0.40 | 0.457 |
| 0.45 | 0.244 |
| 0.50 | 0.000 |

$HPRT^- TK^+$ in the presence of Sendai virus or polyethylene glycol. Fusion takes place in some of the cells, and the mixture is grown in a medium containing hypoxanthine, aminopterin, and thymidine (called **HAT medium**). In the presence of aminopterin, unfused mouse cells ($TK^-$) and unfused human cells ($HPRT^-$) die. Hybrid cells, however, survive because they are $HPRT^+ TK^+$. Eventually, the hybrid cells end up with random numbers of human chromosomes. There is one restriction: All cell lines selected are $TK^+$. This HAT method (using the HAT medium) not only selects for hybrid clones, but also localizes the *TK* gene to human chromosome 17, the one human chromosome found in every successful cell line.

**Table 6.8** **Assignment of the Gene for Blood Coagulating Factor III to Human Chromosome 1 Using Human-Mouse Hybrid Cell Lines**

| Hybrid Cell Line Designation | Tissue/ Factor Score | Human Chromosome Present | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | X |
| *WILI* | − | − | − | − | − | − | − | − | + | − | − | − | − | − | + | − | − | + | − | − | − | + | − | + |
| *WIL6* | − | − | + | − | + | + | + | + | + | − | + | + | − | − | + | − | − | + | − | + | + | + | − | + |
| *WIL7* | − | − | + | + | − | + | + | − | + | − | + | + | − | + | + | − | − | + | + | − | − | + | − | + |
| *WIL14* | + | + | − | + | − | − | − | + | + | − | + | − | + | − | + | + | − | + | − | − | − | − | − | + |
| *SIR3* | + | + | + | + | + | + | + | + | − | + | + | + | + | + | − | − | + | + | + | + | + | + | + | + |
| *SIR8* | + | + | + | + | + | + | − | + | + | + | + | + | + | + | + | + | + | + | + | − | − | + | + | + |
| *SIR11* | − | − | − | − | − | − | − | + | − | − | − | − | − | + | − | − | − | − | − | − | − | + | + | + |
| *REW7* | + | + | + | + | + | + | + | + | + | − | + | + | + | + | + | + | − | + | + | + | + | + | + | + |
| *REW15* | + | + | + | + | + | + | + | + | + | − | + | − | + | + | + | + | − | + | + | + | + | + | + | + |
| *DUA1A* | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | * | − | − | − | − | − | − | − | * |
| *DUA1CsAzF* | − | − | − | − | − | − | − | + | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − |
| *DUA1CsAzH* | − | − | − | − | − | − | − | + | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − |
| *TSL1* | − | − | − | + | + | − | − | − | − | − | + | + | − | + | + | − | + | + | + | − | + | − | − | − |
| *TSL2* | − | − | + | * | − | + | + | − | − | − | + | − | + | − | − | − | − | * | + | − | + | + | − | + |
| *TSL2CsBF* | − | − | − | − | − | + | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − |
| *XTR1* | + | + | − | * | − | + | + | + | + | + | + | + | + | + | + | − | − | + | + | + | + | + | + | + |
| *XTR2* | − | − | − | * | − | + | − | − | + | − | + | − | + | + | − | − | − | − | + | − | + | + | − | * |
| *XTR3BsAgE* | + | + | − | * | − | + | + | + | + | + | + | − | − | + | + | − | − | + | + | + | − | + | − | * |
| *XTR22* | − | − | + | * | + | + | + | − | + | − | + | + | − | − | − | + | − | − | + | + | + | + | + | * |
| *XER9* | − | − | + | − | + | − | − | − | + | − | + | * | + | − | + | − | − | + | + | − | − | + | − | * |
| *XER11* | + | + | − | + | + | − | + | + | + | − | + | * | + | + | − | + | + | + | + | + | + | + | + | * |
| *REX12* | − | − | − | + | − | − | − | + | − | − | − | + | − | − | + | − | − | − | − | − | − | − | + | * |
| *JSR29* | + | + | + | + | + | + | + | * | + | * | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *JVR22* | + | + | + | + | + | + | + | + | + | − | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *JWR22H* | + | * | * | − | + | − | + | − | − | − | + | + | + | − | + | + | − | + | + | − | + | + | − | − |
| *ALR2* | + | + | + | + | + | + | + | + | − | + | + | + | + | + | + | + | + | + | − | + | + | + | + | + |
| *ICL15* | − | − | − | − | − | − | − | − | − | − | − | − | + | − | − | − | − | + | − | − | + | + | − | − |
| *ICL15CsBF* | − | − | − | − | − | − | − | − | − | − | − | − | + | − | − | − | − | − | − | − | + | + | − | − |
| *MH21* | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | + | − | − |
| % Discord[†] | | 0 | 32 | 17 | 24 | 31 | 21 | 21 | 31 | 21 | 24 | 30 | 21 | 21 | 28 | 14 | 24 | 21 | 28 | 17 | 34 | 41 | 21 | 27 |

*Source:* Reprinted with permission from S.D. Carson, et al., "Tissue Factor Gene Localized to Human Chromosome 1 (after 1p21)," *Science*, 229:229–291. Copyright © 1985 American Association for the Advancement of Science.

[*] A translocation in which only part of the chromosome is present.

[†] Discord refers to cases in which the tissue factor score is plus, and the human chromosome is absent, or in which the score is minus and the chromosome is present.

After successful cell hybrids are formed, two particular tests are used to map human genes. A **synteny test** (same linkage group) determines whether two loci are in the same linkage group if the phenotypes of the two loci are either always together or always absent in various hybrid cell lines. An **assignment test** determines which chromosome a particular locus is on by the concordant appearance of the phenotype whenever that particular chromosome is in a cell line, or by the lack of the particular phenotype when a particular chromosome is absent from a cell line. The first autosomal synteny test, performed in 1970, demonstrated that the *B* locus of lactate dehydrogenase ($LDH_B$) was linked to the *B* locus of peptidase ($PEP_B$). (Both enzymes are formed from subunits controlled by two loci each. In addition to the *B* locus, each protein has subunits controlled by an *A* locus.) Later, these loci were shown to reside on chromosome 12.

In another example, a blood-coagulating glycoprotein (a protein-polysaccharide complex) called tissue factor III was localized by assignment tests to chromosome 1. Table 6.8 shows twenty-nine human-mouse hybrid cell lines, or **clones,** the human chromosomes they contain, and their tissue factor score, the results of an assay for the presence of the coagulating factor. (Clones are cells arising from a single ancestor.) It is obvious from table 6.8 that the gene for tissue factor III is on human chromosome 1. Every time human chromosome 1 is present in a cell line, so is tissue factor III. Every time human chromosome 1 is absent, so is the tissue factor (zero discordance or 100% concordance). No other chromosome showed that pattern.

The human map as we know it now (compiled by Victor McKusick at Johns Hopkins University), containing over six thousand assigned loci of over twelve thousand known to exist, is shown in table 6.9 and figure 6.29. At

Victor A. McKusick (1921– ). (Courtesy of Victor A. McKusick.)

**Table 6.9** **Definition of Selected Loci of the Human Chromosome Map (figure 6.29)**

| Locus | Protein Product | Chromosome |
|---|---|---|
| *ABO* | ABO blood group | 9 |
| *AG* | Alpha globin gene family | 16 |
| *ALB* | Albumin | 4 |
| *AMY1* | Amylase, salivary | 1 |
| *AMY2* | Amylase, pancreatic | 1 |
| *BCS* | Breast cancer susceptibility | 16 |
| *C2* | Complement component-2 | 6 |
| *CAT* | Catalase | 11 |
| *CBD* | Color blindness, deutan | X |
| *CBP* | Color blindness, protan | X |
| *CML* | Chronic myeloid leukemia | 22 |
| *DMD* | Duchenne muscular dystrophy | X |
| *FES* | Feline sarcoma virus oncogene | 15 |
| *FY* | Duffy blood group | 1 |
| *GLB1* | Beta-galactosidase-1 | 3 |
| *H1* | Histone-1 | 7 |
| *HBB* | Hemoglobin beta chain | 11 |
| *HEMA* | Classic hemophilia | X |
| *HEXA* | Hexosaminidase A | 15 |
| *HLA* | Human leukocyte antigens | 6 |
| *HP* | Haptoglobin | 16 |
| *HYA* | Y histocompatibility antigen, locus A | Y |
| *IDDM* | Insulin-dependent diabetes mellitus | 6 |
| *IFF* | Interferon, fibroblast | 9 |
| *IGH* | Immunoglobulin heavy-chain gene family | 14 |
| *IGK* | Immunoglobulin kappa-chain gene family | 2 |
| *INS* | Insulin | 11 |
| *LDHA* | Lactate dehydrogenase A | 11 |
| *MDI* | Manic depressive illness | 6 |
| *MHC* | Major histocompatibility complex | 6 |
| *MN* | MN blood group | 4 |
| *MYB* | Avian myeloblastosis virus oncogene | 6 |
| *NHCP1* | Nonhistone chromosomal protein-1 | 7 |
| *NPS1* | Nail-patella syndrome | 9 |
| *PEPA* | Peptidase A | 18 |
| *PVS* | Polio virus sensitivity | 19 |
| *Rh* | Rhesus blood group | 1 |
| *RN5S* | 5S RNA gene(s) | 1 |
| *RNTMI* | Initiator methionine tRNA | 6 |
| *RWS* | Ragweed sensitivity | 6 |
| *S1* | Surface antigen 1 | 11 |
| *SIS* | Simian sarcoma virus oncogene | 22 |
| *STA* | Stature | Y |
| *TF* | Transferrin | 3 |
| *XG* | Xg blood group | X |
| *XRS* | X-ray sensitivity | 13 |

*Note:* A more complete list appears in V. A. McKusick, *Mendelian Inheritance in Man: A Catalog of Human Genes* (Baltimore: Johns Hopkins University Press, 1994).

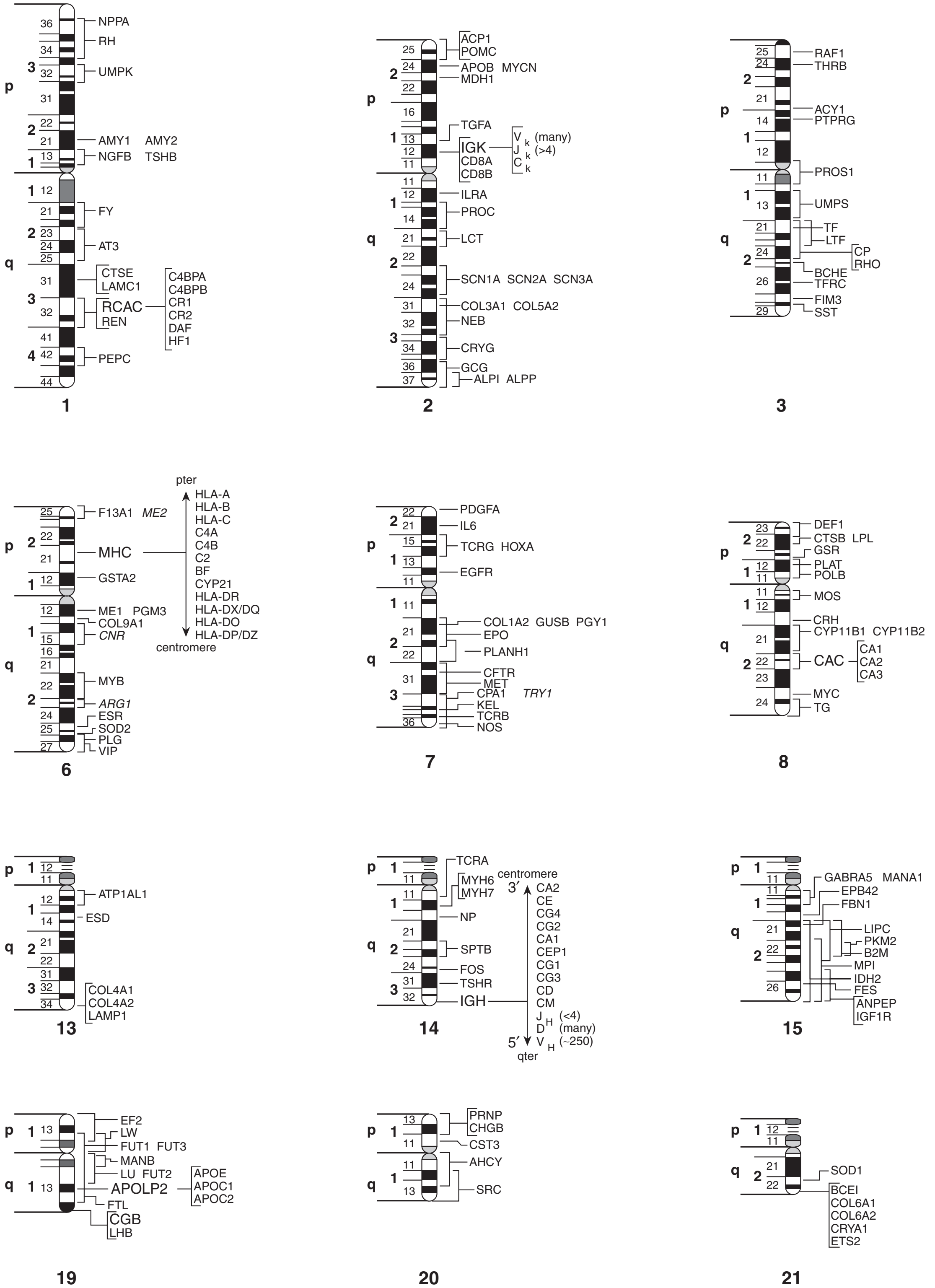
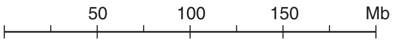

1
NPPA
RH
UMPK
AMY1 AMY2
NGFB TSHB
FY
AT3
CTSE
LAMC1
RCAC
REN
C4BPA
C4BPB
CR1
CR2
DAF
HF1
PEPC
2
ACP1
POMC
APOB MYCN
MDH1
TGFA
IGK
CD8A
CD8B
V k (many)
J k (>4)
C k
ILRA
PROC
LCT
SCN1A SCN2A SCN3A
COL3A1 COL5A2
NEB
CRYG
GCG
ALPI ALPP
3
RAF1
THRB
ACY1
PTPRG
PROS1
UMPS
TF
LTF
CP
RHO
BCHE
TFRC
FIM3
SST
6
F13A1 ME2
MHC
GSTA2
ME1 PGM3
COL9A1
CNR
MYB
ARG1
ESR
SOD2
PLG
VIP
pter
HLA-A
HLA-B
HLA-C
C4A
C4B
C2
BF
CYP21
HLA-DR
HLA-DX/DQ
HLA-DO
HLA-DP/DZ
centromere
7
PDGFA
IL6
TCRG HOXA
EGFR
COL1A2 GUSB PGY1
EPO
PLANH1
CFTR
MET
CPA1 TRY1
KEL
TCRB
NOS
8
DEF1
CTSB LPL
GSR
PLAT
POLB
MOS
CRH
CYP11B1 CYP11B2
CAC
CA1
CA2
CA3
MYC
TG
13
ATP1AL1
ESD
COL4A1
COL4A2
LAMP1
14
TCRA
MYH6
MYH7
NP
SPTB
FOS
TSHR
IGH
centromere
3′
CA2
CE
CG4
CG2
CA1
CEP1
CG1
CG3
CD
CM
J H (<4)
D H (many)
V H (~250)
5′
qter
15
GABRA5 MANA1
EPB42
FBN1
LIPC
PKM2
B2M
MPI
IDH2
FES
ANPEP
IGF1R
19
EF2
LW
FUT1 FUT3
MANB
LU FUT2
APOLP2
APOE
APOC1
APOC2
FTL
CGB
LHB
20
PRNP
CHGB
CST3
AHCY
SRC
21
SOD1
BCEI
COL6A1
COL6A2
CRYA1
ETS2
p
q

**Figure 6.29** Human G-banded chromosomes with their accompanying assigned loci. The *p* and *q* refer to the short and long arms of the chromosomes, respectively. A key to the loci is given in McKusick (1994). (From Victor A. McKusick, *Mendelian inheritance in man,* 11th edition, 1994. Reprinted by permission of Johns Hopkins University Press, Baltimore, MD.)

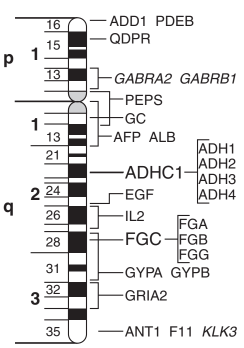
**THE HUMAN GENE MAP**
**(selected 'anchor' loci)**

**A confirmed assignment**.............................. **ENO1**
*A provisional assignment*................................ *DHPR*
**Gene cluster**................................................... **MHC**



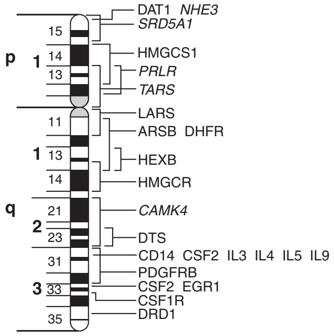
50
100
150
Mb

SCALE
(in megabases)



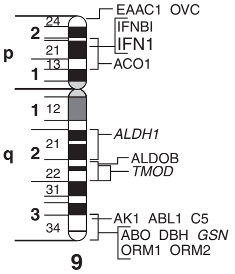
p
1
16
15
13
q
1
13
21
2
24
26
28
31
3
32
35
ADD1 PDEB
QDPR
GABRA2 GABRB1
PEPS
GC
AFP ALB
ADHC1
ADH1
ADH2
ADH3
ADH4
EGF
IL2
FGC
FGA
FGB
FGG
GYPA GYPB
GRIA2
ANT1 F11 KLK3

4



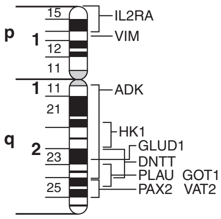
p
1
15
14
13
q
1
11
13
14
2
21
23
31
3
33
35
DAT1 NHE3
SRD5A1
HMGCS1
PRLR
TARS
LARS
ARSB DHFR
HEXB
HMGCR
CAMK4
DTS
CD14 CSF2 IL3 IL4 IL5 IL9
PDGFRB
CSF2 EGR1
CSF1R
DRD1

5



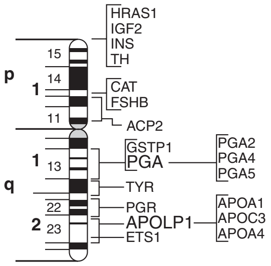
p
2
24
21
1
13
q
1
12
2
21
22
31
3
34
EAAC1 OVC
IFNBI
IFN1
ACO1
ALDH1
ALDOB
TMOD
AK1 ABL1 C5
ABO DBH GSN
ORM1 ORM2

9



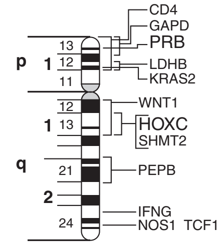
p
1
15
12
11
q
1
11
21
2
23
25
IL2RA
VIM
ADK
HK1
GLUD1
DNTT
PLAU GOT1
PAX2 VAT2

10



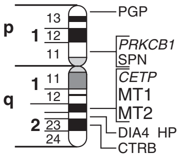
p
1
15
14
11
q
1
13
2
22
23
HRAS1
IGF2
INS
TH
CAT
FSHB
ACP2
GSTP1
PGA
PGA2
PGA4
PGA5
TYR
PGR
APOLP1
APOA1
APOC3
APOA4
ETS1

11



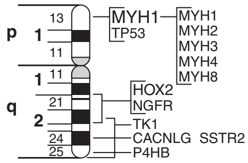
p
1
13
12
11
q
1
12
13
2
21
24
CD4
GAPD
PRB
LDHB
KRAS2
WNT1
HOXC
SHMT2
PEPB
IFNG
NOS1 TCF1

12



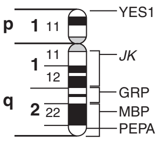
p
1
13
12
11
q
1
11
12
2
23
24
PGP
PRKCB1
SPN
CETP
MT1
MT2
DIA4 HP
CTRB

16



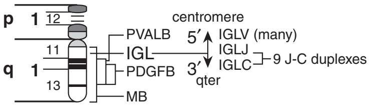
p
1
13
11
q
1
11
2
21
24
25
MYH1
TP53
MYH1
MYH2
MYH3
MYH4
MYH8
HOX2
NGFR
TK1
CACNLG SSTR2
P4HB

17



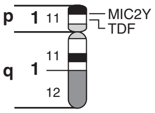
p
1
11
q
1
11
12
2
22
YES1
JK
GRP
MBP
PEPA

18



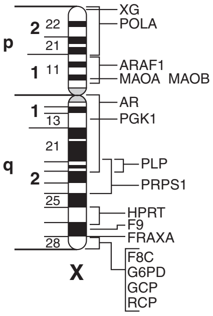
p
1
12
q
1
11
13
PVALB
IGL
PDGFB
MB
centromere
5′
IGLV (many)
IGLJ
IGLC
3′
qter
9 J-C duplexes

22



p
1
11
q
1
11
12
MIC2Y
TDF

Y



p
2
22
21
1
11
q
1
13
21
2
25
28
XG
POLA
ARAF1
MAOA MAOB
AR
PGK1
PLP
PRPS1
HPRT
F9
FRAXA
F8C
G6PD
GCP
RCP

X

present, geneticists studying human chromosomes are hampered not by a lack of techniques but by a lack of marker loci. When a new locus is discovered, it is now relatively easy to assign it to its proper chromosome.

The problem still exists of determining exactly where a particular locus belongs on a chromosome. This is facilitated by developing particular cell lines with broken chromosomes, so that parts are either missing or have moved to other chromosomes. These processes reveal new linkage arrangements and make it possible to determine the region in which a locus is situated on a particular chromosome. In chapter 13, we describe additional techniques used to locate genes on human chromosomes, including a description of the Human Genome Project, the program that sequenced the entire human genome as well as the genomes of other model organisms.

## SUMMARY

**STUDY OBJECTIVE 1:** To learn about analytical techniques for locating the relative positions of genes on chromosomes in diploid eukaryotic organisms 110–122

The principle of independent assortment is violated when loci lie near each other on the same chromosome. Recombination between these loci results from the crossing over of chromosomes during meiosis. The amount of recombination provides a measure of the distance between these loci. One map unit (centimorgan) equals 1% recombinant gametes. Map units can be determined by testcrossing a dihybrid and recording the percentage of recombinant offspring. If three loci are used (a three-point cross), double crossovers will be revealed. A coefficient of coincidence, the ratio of observed to expected double crossovers, can be calculated to determine if one crossover changes the probability that a second one will occur nearby.

A chiasma seen during prophase I of meiosis represents both a physical and a genetic crossing over. This can be demonstrated by using homologous chromosomes with morphological distinctions.

Because of multiple crossovers, the measured percentage recombination underestimates the true map distance, especially for loci relatively far apart; the best map estimates come from summing the distances between closely linked loci. A mapping function can be used to translate observed map distances into more accurate ones.

**STUDY OBJECTIVE 2:** To learn about analytical techniques for locating the relative positions of genes on chromosomes in ascomycete fungi 122–132

Organisms that retain all the products of meiosis lend themselves to chromosome mapping by haploid mapping (tetrad analysis). With unordered spores, such as in yeast, we use

$$\text{map units} = \frac{(1/2)\text{ the number of TT asci} + \text{the number of NPD asci}}{\text{total number or asci}} \times 100$$

Map units between a locus and its centromere in organisms with ordered spores, such as *Neurospora,* can be calculated as

$$\text{map units} = \frac{(1/2)\text{ the number of SDS asci}}{\text{total number of asci}} \times 100$$

Crossing over also occurs during mitosis, but at a much reduced rate. Somatic (mitotic) crossing over can be used to map loci.

**STUDY OBJECTIVE 3:** To learn about analytical techniques for locating the relative positions of genes on human chromosomes 132–140

Human chromosomes can be mapped. Recombination distances can be established by pedigrees, and loci can be attributed to specific chromosomes by synteny and assignment tests in hybrid cell lines.

## SOLVED PROBLEMS

**PROBLEM 1:** A homozygous claret (*ca,* claret eye color), curled (*cu,* upcurved wings), fluted (*fl,* creased wings) fruit fly is crossed with a pure-breeding wild-type fly. The $F_1$ females are testcrossed with the following results:

| | |
|---|---|
| fluted | 4 |
| claret | 173 |
| curled | 26 |
| fluted, claret | 24 |
| fluted, curled | 167 |
| claret, curled | 6 |
| fluted, claret, curled | 298 |
| wild-type | 302 |

**a.** Are the loci linked?
**b.** If so, give the gene order, map distances, and coefficient of coincidence.

***Answer:*** The pattern of numbers among the eight offspring classes is the pattern we are used to seeing for linkage of three loci. We can tell from the two groups in largest numbers (the nonrecombinants—fluted, claret, curled and wild-type) that the alleles are in the coupling (*cis*) arrangement. If we compare either of the nonrecombinant classes with either of the double crossover classes (fluted and claret, curled), we see that the fluted locus is in the center. For example, compare fluted, a double crossover offspring, with the wild-type, a nonrecombinant; clearly, fluted has the odd pattern. Thus the trihybrid female parent had the following arrangement of alleles:

$$\frac{ca\ fl\ cu}{ca^+\ fl^+\ cu^+}$$

A crossover in the *ca–fl* region produces claret and fluted, curled offspring, and a crossover in the *fl–cu* region produces fluted, claret and curled offspring. Counting the crossovers in each region, including the double crossovers in each, and then converting to percentages, yields a claret-to-fluted distance of 35.0 map units (173 + 167 + 6 + 4) and a fluted-to-curled distance of 6.0 map units (26 + 24 + 6 + 4). We expect 0.35 × 0.06 × 1,000 = 21 double crossovers, but we observed only 6 + 4 = 10. Thus, the coefficient of coincidence is 10/21 = 0.48.

**PROBLEM 2:** The *ad5* locus in *Neurospora* is a gene for an enzyme in the synthesis pathway for the DNA base adenine. A wild-type strain ($ad5^+$) is crossed with an adenine-requiring strain, $ad5^-$. The diploid undergoes meiosis, and one hundred asci are scored for their segregation patterns with the following results:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $ad5^+$ | $ad5^+$ | $ad5^+$ | $ad5^+$ | $ad5^-$ | $ad5^-$ | $ad5^-$ | $ad5^-$ | 40 |
| $ad5^-$ | $ad5^-$ | $ad5^-$ | $ad5^-$ | $ad5^+$ | $ad5^+$ | $ad5^+$ | $ad5^+$ | 46 |
| $ad5^+$ | $ad5^+$ | $ad5^-$ | $ad5^-$ | $ad5^-$ | $ad5^-$ | $ad5^+$ | $ad5^+$ | 5 |
| $ad5^-$ | $ad5^-$ | $ad5^+$ | $ad5^+$ | $ad5^+$ | $ad5^+$ | $ad5^-$ | $ad5^-$ | 3 |
| $ad5^-$ | $ad5^-$ | $ad5^+$ | $ad5^+$ | $ad5^-$ | $ad5^-$ | $ad5^+$ | $ad5^+$ | 4 |
| $ad5^+$ | $ad5^+$ | $ad5^-$ | $ad5^-$ | $ad5^+$ | $ad5^+$ | $ad5^-$ | $ad5^-$ | 2 |

What can you say about the linkage arrangements at this locus?

***Answer:*** You can see that 14 (5 + 3 + 4 + 2) asci are of the second-division segregation type (SDS) and 86 (40 + 46) are of the first-division segregation type (FDS). To map the distance of the locus to its centromere, we divide the percentage of SDS types by 2: 14/100 = 14%; divided by 2 is 7%. Thus, the *ad5* locus is 7 map units from its centromere.

**PROBLEM 3:** In yeast, the *his5* locus is a gene for an enzyme in the synthesis pathway for the amino acid histidine, and the *lys11* locus is a gene for an enzyme in the synthesis pathway for the amino acid lysine. A haploid wild-type strain ($his5^+\ lys11^+$) is crossed with the double mutant ($his5^-\ lys11^-$). The diploid is allowed to undergo meiosis, and 100 asci are scored with the following results:

| | | |
|---|---|---|
| $his5^+\ lys11^+$ | $his5^+\ lys11^+$ | $his5^+\ lys11^-$ |
| $his5^+\ lys11^+$ | $his5^-\ lys11^-$ | $his5^+\ lys11^-$ |
| $his5^-\ lys11^-$ | $his5^-\ lys11^+$ | $his5^-\ lys11^+$ |
| $his5^-\ lys11^-$ | $his5^+\ lys11^-$ | $his5^-\ lys11^+$ |
| 62 | 30 | 8 |

What is the linkage arrangement of these loci?

***Answer:*** Of the 100 asci analyzed, 62 were parental ditypes (PD), 30 were tetratypes (TT), and 8 were nonparental ditypes (NPD). To map the distance between the two loci, we take the percentage of NPD (8%) plus half the percentage of TT (1/2 of 30 = 15%) = 23% or 23 centimorgans between loci.

**PROBLEM 4:** A particular human enzyme is present only in clone B. The human chromosomes present in clones A, B, and C appear as pluses in the following table. Determine the probable chromosomal location of the gene for the enzyme.

| | **Human Chromosome** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Clone | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* |
| A | + | + | + | + | − | − | − | − |
| B | + | + | − | − | + | + | − | − |
| C | + | − | + | − | + | − | + | − |

***Answer:*** If a gene is located on a chromosome, the gene must be present in the clones with the chromosome (+). Chromosomes 1, 2, 5, 6 are present in B. If the gene in question were located on chromosome 1, the enzyme should have been present in all three clones. A similar argument holds for chromosome 2, in which the enzyme should have been present in clones A and B, and so on for the rest of the chromosomes. The only chromosome that is unique to clone B is 6. Therefore, the gene is located on chromosome 6.

# EXERCISES AND PROBLEMS*

## DIPLOID MAPPING

**1.** A homozygous groucho fly (*gro,* bristles clumped above the eyes) is crossed with a homozygous rough fly (*ro,* eye abnormality). The $F_1$ females are testcrossed, producing these offspring:

| | |
|---|---|
| groucho | 518 |
| rough | 471 |
| groucho, rough | 6 |
| wild-type | 5 |
| | 1,000 |

**a.** What is the linkage arrangement of these loci?

**b.** What offspring would result if the $F_1$ dihybrids were crossed among themselves instead of being testcrossed?

**2.** A female fruit fly with abnormal eyes (*abe*) of a brown color (*bis,* bistre) is crossed with a wild-type male. Her sons have abnormal, brown eyes; her daughters are of the wild-type. When these $F_1$ flies are crossed among themselves, the following offspring are produced:

| | **Sons** | **Daughters** |
|---|---|---|
| abnormal, brown | 219 | 197 |
| abnormal | 43 | 45 |
| brown | 37 | 35 |
| wild-type | 201 | 223 |

What is the linkage arrangement of these loci?

**3.** In *Drosophila,* the loci inflated (*if*, small, inflated wings) and warty *(wa*, abnormal eyes) are about 10 map units apart on the X chromosome. Construct a data set that would allow you to determine this linkage arrangement. What differences would be involved if the loci were located on an autosome?

**4.** A geneticist crossed female fruit flies that were heterozygous at three electrophoretic loci, each with fast and slow alleles, with males homozygous for the slow alleles. The three loci were *got1* (glutamate oxaloacetate transaminase-1), *amy* (alpha-amylase), and *sdh* (succinate dehydrogenase). The first 1,000 offspring isolated had the following genotypes:

| | | |
|---|---|---|
| Class 1 | $got^s\,got^s\,amy^s\,amy^s\,sdh^s\,sdh^s$ | 441 |
| Class 2 | $got^f\,got^s\,amy^f\,amy^s\,sdh^f\,sdh^s$ | 421 |
| Class 3 | $got^f\,got^s\,amy^s\,amy^s\,sdh^s\,sdh^s$ | 11 |
| Class 4 | $got^s\,got^s\,amy^f\,amy^s\,sdh^f\,sdh^s$ | 14 |
| Class 5 | $got^f\,got^s\,amy^f\,amy^s\,sdh^s\,sdh^s$ | 58 |
| Class 6 | $got^s\,got^s\,amy^s\,amy^s\,sdh^f\,sdh^s$ | 53 |
| Class 7 | $got^f\,got^s\,amy^s\,amy^s\,sdh^f\,sdh^s$ | 1 |
| Class 8 | $got^s\,got^s\,amy^f\,amy^s\,sdh^s\,sdh^s$ | 1 |

What are the linkage arrangements of these three loci, including map units? If the three loci are linked, what is the coefficient of coincidence?

**5.** The following three recessive markers are known in lab mice: *h*, hotfoot; *o*, obese; and *wa*, waved. A trihybrid of unknown origin is testcrossed, producing the following offspring:

| | |
|---|---|
| hotfoot, obese, waved | 357 |
| hotfoot, obese | 74 |
| waved | 66 |
| obese | 79 |
| wild-type | 343 |
| hotfoot, waved | 61 |
| obese, waved | 11 |
| hotfoot | 9 |
| | 1,000 |

**a.** If the genes are linked, determine the relative order and the map distances between them.

**b.** What was the *cis-trans* allele arrangement in the trihybrid parent?

**c.** Is there any crossover interference? If yes, how much?

**6.** The following three recessive genes are found in corn: *bt1*, brittle endosperm; *gl17*, glossy leaf; *rgdl,* ragged seedling. A trihybrid of unknown origin is testcrossed, producing the following offspring:

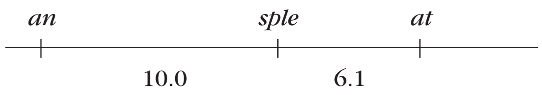
| | |
|---|---|
| brittle, glossy, ragged | 236 |
| brittle, glossy | 241 |
| ragged | 219 |
| glossy | 23 |
| wild-type | 224 |
| brittle, ragged | 17 |
| glossy, ragged | 21 |
| brittle | 19 |
| | 1,000 |

**a.** If the genes are linked, determine the relative order and map distances.

**b.** Reconstruct the chromosomes of the trihybrid.

**c.** Is there any crossover interference? If yes, how much?

**7.** In *Drosophila,* ancon (*an*, legs and wings short), spiny legs (*sple*, irregular leg hairs), and arctus oculus (*at*, small narrow eyes) have the following linkage arrangement on chromosome 3:



---

*Answers to selected exercises and problems are on page A-6.

- **a.** Devise a data set with no crossover interference that would yield these map units.
- **b.** What data would yield the same map units but with a coefficient of coincidence of 0.60?

**8.** Ancon (*an*) and spiny legs (*sple*), from problem 7, are 10 map units apart on chromosome 3. Notchy (*ny,* wing tips nicked) is on the X chromosome (chromosome 1). Create a data set that would result if you were making crosses to determine the linkage arrangement of these three loci. How would you know that the notchy locus is on the X chromosome?

**9.** In the house mouse, the autosomal alleles Trembling and Rex (short hair) are dominant to not trembling (normal) and long hair, respectively. Heterozygous Trembling, Rex females were crossed with normal, long-haired males and yielded the following offspring:

| | |
|---|---|
| Trembling, Rex | 42 |
| Trembling, long-haired | 105 |
| normal, Rex | 109 |
| normal, long-haired | 44 |

- **a.** Are the two genes linked? How do you know?
- **b.** In the heterozygous females, were Trembling and Rex in *cis* or *trans* position? Explain.
- **c.** Calculate the percent recombination between the two genes.

**10.** In corn, a trihybrid Tunicate (*T*), Glossy (*G*), Liguled (*L*) plant was crossed with a nontunicate, nonglossy, liguleless plant, producing the following offspring:

| | |
|---|---|
| Tunicate, liguleless, Glossy | 58 |
| Tunicate, liguleless, nonglossy | 15 |
| Tunicate, Liguled, Glossy | 55 |
| Tunicate, Liguled, nonglossy | 13 |
| nontunicate, Liguled, Glossy | 16 |
| nontunicate, Liguled, nonglossy | 53 |
| nontunicate, liguleless, Glossy | 14 |
| nontunicate, liguleless, nonglossy | 59 |

- **a.** Determine which genes are linked.
- **b.** Determine the genotype of the heterozygote; be sure to indicate which alleles are on which chromosome.
- **c.** Calculate the map distances between the linked genes.

**11.** In *Drosophila*, kidney-shaped eye (*k*), cardinal eye (*cd*), and ebony body (*e*) are three recessive genes. If homozygous kidney, cardinal females are crossed with homozygous ebony males, the $F_1$ offspring are all wild-type. If heterozygous $F_1$ females are mated with kidney, cardinal, ebony males, the following 2,000 progeny appear:

| | |
|---|---|
| 880 | kidney, cardinal |
| 887 | ebony |
| 64 | kidney, ebony |
| 67 | cardinal |
| 49 | kidney |
| 46 | ebony, cardinal |
| 3 | kidney, ebony, cardinal |
| 4 | wild-type |

- **a.** Determine the chromosomal composition of the $F_1$ females.
- **b.** Derive a map of the three genes.

**12.** Following is a partial map of the third chromosome in *Drosophila.*

19.2 javelin bristles (*jv*)
43.2 thread arista (*th*)
66.2 Delta veins (*Dl*)
70.7 ebony body (*e*)

- **a.** If flies heterozygous in *cis* position for javelin and ebony are mated among themselves, what phenotypic ratio do you expect in the progeny?
- **b.** A true-breeding thread, ebony fly is crossed with a true-breeding Delta fly. An $F_1$ female is testcrossed to a thread, ebony male. Predict the expected progeny and their frequencies for this cross. Assume no interference.
- **c.** Repeat *b,* but assume a coefficient of coincidence of 0.4.

**13.** Suppose that you have determined the order of three genes to be *a*, *c*, *b*, and that by doing two-point crosses you have determined map distances as $a$–$c$ = 10 and $c$–$b$ = 5. If interference is –1.5, and the three-point cross is

$$\frac{A\ C\ B}{a\ c\ b} \times \frac{a\ c\ b}{a\ c\ b}$$

what frequency of double crossovers do you expect?

### HAPLOID MAPPING (TETRAD ANALYSIS)

**14.** Given the following cross in *Neurospora:* $ab \times a^+b^+$, construct results showing that crossing over occurs in two of the four chromatids of a tetrad at meiosis. What would the results be if crossing over occurred during interphase before each chromosome became two chromatids? if each crossover event involved three or four chromatids?

**15.** A strain of yeast requiring both tyrosine ($tyr^-$) and arginine ($arg^-$) is crossed to the wild-type. After meiosis, the following ten asci are dissected. Classify each ascus as to segregational type (PD, NPD, TT). What is the linkage relationship between these two loci?

| | | | | |
|---|---|---|---|---|
| 1 | $arg^- \; tyr^-$ | $arg^+ \; tyr^+$ | $arg^+ \; tyr^+$ | $arg^- \; tyr^-$ |
| 2 | $arg^+ \; tyr^+$ | $arg^+ \; tyr^+$ | $arg^- \; tyr^-$ | $arg^- \; tyr^-$ |
| 3 | $arg^- \; tyr^+$ | $arg^- \; tyr^+$ | $arg^+ \; tyr^-$ | $arg^+ \; tyr^-$ |
| 4 | $arg^- \; tyr^-$ | $arg^- \; tyr^-$ | $arg^+ \; tyr^+$ | $arg^+ \; tyr^+$ |
| 5 | $arg^- \; tyr^-$ | $arg^- \; tyr^+$ | $arg^+ \; tyr^-$ | $arg^+ \; tyr^+$ |
| 6 | $arg^+ \; tyr^+$ | $arg^+ \; tyr^+$ | $arg^- \; tyr^-$ | $arg^- \; tyr^-$ |
| 7 | $arg^- \; tyr^-$ | $arg^+ \; tyr^+$ | $arg^- \; tyr^+$ | $arg^+ \; tyr^-$ |
| 8 | $arg^+ \; tyr^+$ | $arg^+ \; tyr^+$ | $arg^- \; tyr^-$ | $arg^- \; tyr^-$ |
| 9 | $arg^+ \; tyr^+$ | $arg^- \; tyr^-$ | $arg^- \; tyr^-$ | $arg^+ \; tyr^+$ |
| 10 | $arg^- \; tyr^-$ | $arg^+ \; tyr^+$ | $arg^+ \; tyr^+$ | $arg^- \; tyr^-$ |

**16.** A certain haploid strain of yeast was deficient for the synthesis of the amino acids tryptophan ($try^-$) and methionine ($met^-$). It was crossed to the wild-type, and meiosis occurred. One dozen asci were analyzed for their tryptophan and methionine requirements. The following results, with the inevitable lost spores, were obtained:

| | | | | |
|---|---|---|---|---|
| 1 | $try^- \; met^-$ | ? | ? | $try^- \; met^-$ |
| 2 | ? | $try^- \; met^-$ | $try^+ \; met^+$ | $try^+ \; met^+$ |
| 3 | $try^- \; met^+$ | $try^- \; met^-$ | $try^+ \; met^-$ | $try^+ \; met^+$ |
| 4 | $try^- \; met^-$ | $try^+ \; met^+$ | ? | $try^+ \; met^-$ |
| 5 | $try^- \; met^+$ | ? | ? | $try^+ \; met^-$ |
| 6 | $try^+ \; met^+$ | $try^+ \; met^+$ | $try^- \; met^-$ | $try^- \; met^-$ |
| 7 | $try^+ \; met^+$ | $try^+ \; met^-$ | ? | $try^- \; met^-$ |
| 8 | $try^+ \; met^+$ | $try^- \; met^-$ | ? | $try^+ \; met^+$ |
| 9 | $try^- \; met^+$ | $try^+ \; met^-$ | $try^- \; met^+$ | $try^+ \; met^-$ |
| 10 | $try^- \; met^-$ | $try^+ \; met^+$ | $try^- \; met^-$ | $try^+ \; met^+$ |
| 11 | $try^+ \; met^+$ | $try^+ \; met^+$ | ? | ? |
| 12 | ? | $try^+ \; met^-$ | ? | $try^- \; met^+$ |

- **a.** Classify each ascus as to segregational type (note that some asci may not be classifiable).
- **b.** Are the genes linked?
- **c.** If so, how far apart are they?

**17.** In *Neurospora*, a haploid strain requiring arginine ($arg^-$) is crossed with the wild-type ($arg^+$). Meiosis occurs, and ten asci are dissected with the following results. Map the *arg* locus.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | $arg^+$ | $arg^+$ | $arg^-$ | $arg^-$ | $arg^+$ | $arg^+$ | $arg^-$ | $arg^-$ |
| 2 | $arg^-$ | $arg^-$ | $arg^+$ | $arg^+$ | $arg^-$ | $arg^-$ | $arg^+$ | $arg^+$ |
| 3 | $arg^+$ | $arg^+$ | $arg^+$ | $arg^+$ | $arg^-$ | $arg^-$ | $arg^-$ | $arg^-$ |
| 4 | $arg^+$ | $arg^+$ | $arg^+$ | $arg^+$ | $arg^-$ | $arg^-$ | $arg^-$ | $arg^-$ |
| 5 | $arg^-$ | $arg^-$ | $arg^-$ | $arg^-$ | $arg^+$ | $arg^+$ | $arg^+$ | $arg^+$ |
| 6 | $arg^+$ | $arg^+$ | $arg^-$ | $arg^-$ | $arg^-$ | $arg^-$ | $arg^+$ | $arg^+$ |
| 7 | $arg^-$ | $arg^-$ | $arg^+$ | $arg^+$ | $arg^+$ | $arg^+$ | $arg^-$ | $arg^-$ |
| 8 | $arg^+$ | $arg^+$ | $arg^+$ | $arg^+$ | $arg^-$ | $arg^-$ | $arg^-$ | $arg^-$ |
| 9 | $arg^-$ | $arg^-$ | $arg^+$ | $arg^+$ | $arg^+$ | $arg^+$ | $arg^-$ | $arg^-$ |
| 10 | $arg^-$ | $arg^-$ | $arg^-$ | $arg^-$ | $arg^+$ | $arg^+$ | $arg^+$ | $arg^+$ |

**18.** A haploid strain of *Neurospora* with fuzzy colony morphology ($f$) was crossed with the wild-type ($f^+$). Twelve asci were scored. The following results, with the inevitable lost spores were obtained:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | ? | $f$ | $f$ | ? | ? | $f^+$ | $f^+$ | $f^+$ |
| 2 | $f$ | $f$ | $f^+$ | $f^+$ | $f^+$ | $f^+$ | $f$ | $f$ |
| 3 | $f$ | ? | ? | ? | $f^+$ | ? | ? | ? |
| 4 | $f^+$ | ? | ? | ? | $f$ | $f$ | $f$ | $f$ |
| 5 | $f$ | $f$ | ? | ? | ? | $f^+$ | ? | $f^+$ |
| 6 | ? | $f$ | $f$ | ? | ? | ? | ? | ? |
| 7 | $f^+$ | $f^+$ | $f$ | $f$ | $f$ | $f$ | $f^+$ | $f^+$ |
| 8 | $f$ | $f$ | $f$ | ? | ? | $f^+$ | $f^+$ | $f^+$ |
| 9 | $f^+$ | ? | ? | ? | ? | $f$ | $f$ | ? |
| 10 | $f$ | $f$ | $f^+$ | $f^+$ | $f$ | $f$ | $f^+$ | $f^+$ |
| 11 | $f$ | $f$ | $f$ | $f$ | $f^+$ | $f^+$ | $f^+$ | $f^+$ |
| 12 | $f$ | $f$ | ? | ? | ? | ? | $f^+$ | $f^+$ |

- **a.** Classify each ascus as to segregational type, and note which asci cannot be classified.
- **b.** Map the chromosome containing the $f$ locus with all the relevant measurements.

**19.** Draw ten of the remaining twenty-eight ascus patterns not included in table 6.6. To which of the seven major categories of table 6.7 does each belong?

**20.** In yeast, the $a$ and $b$ loci are 12 map units apart. Construct a data set to demonstrate this.

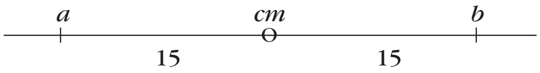
**21.** In *Neurospora*, the $a$ locus is 12 map units from its centromere. Construct a data set to show this.

**22.** An *ab Neurospora* was crossed with an $a^+b^+$ form. Meiosis occurred, and 1,000 asci were dissected. Using the classes of table 6.7, the following data resulted:

| | | | |
|---|---|---|---|
| Class 1 | 700 | Class 5 | 5 |
| Class 2 | 0 | Class 6 | 5 |
| Class 3 | 190 | Class 7 | 10 |
| Class 4 | 90 | | |

What is the linkage arrangement of these loci?

**23.** Given the following linkage arrangement in *Neurospora*, construct a data set similar to that in table 6.7 that is consistent with it (*cm* is centromere).



**24.** Determine crossover events that led to each of the seven classes in table 6.7.

**25.** In *Neurospora,* a cross is made between $ab^+$ and $a^+b$ individuals. The following one hundred ordered tetrads are obtained:

| Spores | I | II | III | IV | V | VI | VII | VIII |
|---|---|---|---|---|---|---|---|---|
| 1, 2 | $a^+b$ | $a^+b$ | $a^+b$ | $a^+b^+$ | $a^+b^+$ | $a^+b$ | $a^+b$ | $ab^+$ |
| 3, 4 | $a^+b$ | $a^+b^+$ | $a^+b^+$ | $a^+b$ | $a^+b$ | $ab^+$ | $ab^+$ | $a^+b$ |
| 5, 6 | $ab^+$ | $ab$ | $ab^+$ | $ab$ | $ab^+$ | $a^+b$ | $ab^+$ | $a^+b$ |
| 7, 8 | $ab^+$ | $ab^+$ | $ab$ | $ab^+$ | $ab$ | $ab^+$ | $a^+b$ | $ab^+$ |
| | 85 | 2 | 3 | 2 | 3 | 3 | 1 | 1 |

- **a.** Are genes $a$ and $b$ linked? How do you know?
- **b.** Calculate the gene-to-centromere distances for $a$ and $b$.

26. *Neurospora* has four genes—*a*, *b*, *c*, and *d*—that control four different phenotypes. Your job is to map these genes by performing pairwise crosses. You obtain the following ordered tetrads:

| | $ab^+ \times a^+b$ | | | | $bc^+ \times b^+c$ | | |
|---|---|---|---|---|---|---|---|
| **Spores** | **I** | **II** | **III** | **Spores** | **I** | **II** | **III** |
| 1, 2 | $ab^+$ | $ab$ | $ab^+$ | 1, 2 | $bc^+$ | $b^+c^+$ | $b^+c$ |
| 3, 4 | $ab^+$ | $ab$ | $a^+b^+$ | 3, 4 | $bc^+$ | $b^+c^+$ | $b^+c^+$ |
| 5, 6 | $a^+b$ | $a^+b^+$ | $a^+b$ | 5, 6 | $b^+c$ | $bc$ | $bc$ |
| 7, 8 | $a^+b$ | $a^+b^+$ | $ab$ | 7, 8 | $b^+c$ | $bc$ | $bc^+$ |
| | 45 | 43 | 12 | | 70 | 4 | 26 |

| | $cd^+ \times c^+d$ | | | | | | |
|---|---|---|---|---|---|---|---|
| **Spores** | **I** | **II** | **III** | **IV** | **V** | **VI** | **VII** |
| 1, 2 | $cd^+$ | $cd$ | $cd$ | $cd$ | $cd^+$ | $cd$ | $cd^+$ |
| 3, 4 | $cd^+$ | $cd$ | $cd^+$ | $c^+d$ | $c^+d$ | $c^+d^+$ | $c^+d$ |
| 5, 6 | $c^+d$ | $c^+d^+$ | $c^+d^+$ | $c^+d^+$ | $c^+d$ | $c^+d^+$ | $c^+d^+$ |
| 7, 8 | $c^+d$ | $c^+d^+$ | $c^+d$ | $cd^+$ | $cd^+$ | $cd$ | $cd$ |
| | 42 | 2 | 30 | 15 | 5 | 1 | 5 |

- **a.** Calculate the gene-to-centromere distances.
- **b.** Which genes are linked? Explain.
- **c.** Derive a complete map for all four genes.

27. You have isolated a new fungus and have obtained a strain that requires both arginine ($arg^-$) and adenine ($ad^-$). You cross these two strains and collect four hundred random spores that you plate on minimal medium. If twenty-five spores grow, what is the distance between these two genes?

28. Three distinct genes, *pab*, *pk*, and *ad*, were scored in a cross of *Neurospora*. From the cross $pab\ pk^+\ ad^+ \times pab^+\ pk\ ad$, the following ordered tetrads were recovered:

| **Spores** | **I** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
|---|---|---|---|---|---|---|---|---|
| 1, 2 | $pab\ pk^+\ ad^+$ | $pab\ pk^+\ ad^+$ | $pab\ pk^+\ ad^+$ | $pab\ pk^+\ ad^+$ | $pab\ pk^+\ ad^+$ | $pab\ pk^+\ ad^+$ | $pab\ pk^+\ ad$ | $pab\ pk^+\ ad$ |
| 3, 4 | $pab\ pk^+\ ad^+$ | $pab^+pk\ ad$ | $pab\ pk\ ad$ | $pab\ pk^+ad$ | $pab^+\ pk\ ad$ | $pab^+\ pk\ ad$ | $pab^+\ pk\ ad$ | $pab^+pk\ ad^+$ |
| 5, 6 | $pab^+pk\ ad$ | $pab\ pk^+ad^+$ | $pab^+pk^+ad^+$ | $pab^+pk\ ad^+$ | $pab\ pk\ ad$ | $pab\ pk^+ad$ | $pab^+pk\ ad^+$ | $pab\ pk^+ad^+$ |
| 7, 8 | $pab^+pk\ ad$ | $pab^+pk\ ad$ | $pab^+pk\ ad$ | $pab^+pk\ ad$ | $pab^+pk^+\ ad^+$ | $pab^+pk\ ad^+$ | $pab\ pk^+ad^+$ | $pab^+pk\ ad$ |
| | 34 | 35 | 9 | 7 | 2 | 2 | 1 | 3 |

Based on the data, construct a map of the three genes. Be sure to indicate centromeres.

### HUMAN CHROMOSOMAL MAPS

29. The Duffy blood group with alleles $FY^a$ and $FY^b$ was localized to chromosome 1 in human beings when an "uncoiled" chromosome was associated with it. Construct a pedigree that would verify this.

**30.** What pattern of scores would you expect to get, using the hybrid clones in table 6.8, for a locus on human chromosome 6? 14? X?

**31.** A man with X-linked color blindness and X-linked Fabry disease (alpha-galactosidase-A deficiency) mates with a normal woman and has a normal daughter. This daughter then mates with a normal man and produces ten sons (as well as eight normal daughters). Of the sons, five are normal, three are like their grandfather, one is only color-blind, and one has Fabry disease. From these data, what can you say about the relationship of these two X-linked loci?

**32.** In people, the ABO system ($I^A$, $I^B$, *i* alleles) is linked to the aldolase-B locus (*ALDOB*), a gene that functions in the liver. Deficiency, which is recessive, results in fructose intolerance. A man with blood type AB has a fructose-intolerant, type B father and a normal, type AB mother. He and a woman with blood type O and fructose intolerance have ten children. Five are type A and normal, three are fructose intolerant and type B, and two are type A and intolerant to fructose. Draw a pedigree of this family and determine the map distances involved. (Calculate a *lod* score to determine the most likely recombination frequency between the loci.)

**33.** Hemophilia and color-blindness are X-linked recessive traits. A normal woman whose mother was color-blind and whose father was a hemophiliac mates with a normal man whose father was color-blind. They have the following children:

4 normal daughters
1 normal son
2 color-blind sons
2 hemophiliac sons
1 color-blind, hemophiliac son

Estimate the distance between the two genes.

**34.** The results of an analysis of five human-mouse hybrids for five enzymes are given in table along with the human chromosomal content of each clone (+ = enzyme or chromosome present; − = absent). Deduce which chromosome carries which gene.

| | Clone | | | | |
|---|---|---|---|---|---|
| **Human Enzyme** | **A** | **B** | **C** | **D** | **E** |
| glutathione reductase | + | + | − | − | − |
| malate dehydrogenase | − | + | − | − | − |
| adenosine deaminase | − | + | − | + | + |
| galactokinase | − | + | + | − | − |
| hexosaminidase | + | − | − | + | − |

| | Human Chromosome | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** |
| Clone A | − | − | − | − | + | + | + | + | − | + | − | − | − | − | + | + | − | − | − | − | + | + |
| Clone B | + | + | − | + | − | − | − | + | − | − | + | − | − | + | − | − | + | − | − | + | − | − |
| Clone C | − | − | − | + | − | − | + | − | − | + | − | + | + | + | + | − | + | − | + | − | − | + |
| Clone D | + | − | + | − | + | − | − | − | − | + | − | − | − | + | + | − | − | + | + | + | + | − |
| Clone E | − | − | − | + | − | − | − | − | + | + | + | + | − | + | − | + | − | + | − | + | + | + |

**35.** You have selected three mouse-human hybrid clones and analyzed them for the presence of human chromosomes. You then analyze each clone for the presence or absence of particular human enzymes (+ = presence of human chromosome or enzyme activity). Based on the following results indicate the probable chromosomal location for each enzyme.

| | **Human Chromosomes** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Clone** | **3** | **7** | **9** | **11** | **15** | **18** | **20** |
| X | − | + | − | + | + | − | + |
| Y | + | + | − | + | − | + | − |
| Z | − | + | + | − | − | + | + |

| | **Enzyme** | | | | |
|---|---|---|---|---|---|
| **Clone** | **A** | **B** | **C** | **D** | **E** |
| X | + | + | − | − | + |
| Y | + | − | + | + | + |
| Z | − | − | + | − | + |

**36.** Three mouse-human cell lines were scored for the presence (+) or absence (−) of human chromosomes, with the results as follows:

| | **Human Chromosomes** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Clone** | **1** | **2** | **3** | **4** | **5** | **14** | **15** | **18** |
| A | + | + | + | + | − | − | − | − |
| B | + | + | − | − | + | + | − | − |
| C | + | − | + | − | + | − | + | − |

If a particular gene is located on chromosome 3, which clones should be positive for the enzyme from that gene?

## CRITICAL THINKING QUESTIONS

1. Do three-point crosses in fruit flies capture all the multiple crossovers in a region?
2. If 4% of all tetrads have a single crossover between two loci: (a) What is the map distance between these loci if these are fruit flies? (b) What is the proportion of second-division segregants if these are *Neurospora?* (c) What is the proportion of nonparental ditypes if these are yeast?

*Suggested Readings for chapter 6 are on page B-3.*

# 7

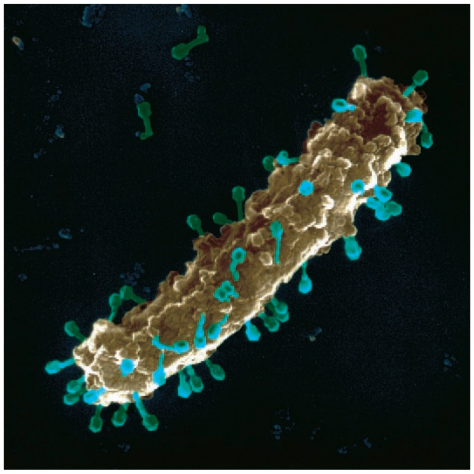
# LINKAGE AND MAPPING IN PROKARYOTES AND BACTERIAL VIRUSES

## STUDY OBJECTIVES

1. To define bacteria and bacterial viruses and learn about methods of studying them 149
2. To study life cycles and sexual processes in bacteria and bacteriophages 154, 163
3. To make use of the sexual processes of bacteria and their viruses to map their chromosomes 155, 166

## STUDY OUTLINE

**Bacteria and Bacterial Viruses in Genetic Research** 149
**Techniques of Cultivation** 150
**Bacterial Phenotypes** 151
  Colony Morphology 151
  Nutritional Requirements 151
  Resistance and Sensitivity 153
**Viral Phenotypes** 154
**Sexual Processes in Bacteria and Bacteriophages** 154
  Transformation 154
  Conjugation 157
**Life Cycles of Bacteriophages** 163
  Recombination 163
  Lysogeny 165
**Transduction** 165
  Specialized Transduction 165
  Generalized Transduction 166
  Mapping with Transduction 166
*Summary* 171
*Solved Problems* 172
*Exercises and Problems* 172
*Critical Thinking Questions* 176



Scanning electron micrograph (color enhanced) of an *Escherichia coli* bacterium with adsorbed T-family bacteriophages (36,000x). (© Oliver Meckes/MPI-Tubingen/Photo Researchers.)

All organisms and viruses have genes located sequentially in their genetic material; and almost all can undergo recombination between homologous (equivalent) pieces of genetic material. Because recombination can occur, it is possible to map, by analytical methods, the locations and sequence of genes along the chromosomes of all organisms and almost all viruses. In this chapter, the viruses we look at are those that attack bacteria. Through work with bacteria and viruses, we have entered the modern era of molecular genetics, the subject of the next section of this book.

Bacteria (including the cyanobacteria, the blue-green algae) are prokaryotes. The prokaryotes also include the **archaea,** or archaebacteria, a kingdom recognized in 1980. These highly specialized organisms (previously classified as bacteria), along with the bacteria and eukaryotes, make up the three domains of life on Earth.

The true bacteria can be classified according to shape: a spherical bacterium is called a **coccus;** a rod-shaped bacterium is called a **bacillus;** and a spiral bacterium is called a **spirillum.** Prokaryotes do not undergo mitosis or meiosis but simply divide in two after their chromosome (usually only one), most often a circle of DNA, has replicated (see chapter 9). Bacterial viruses do not even divide; they are mass-produced within a host cell.

## BACTERIA AND BACTERIAL VIRUSES IN GENETIC RESEARCH

Several properties of bacteria and viruses make them especially suitable for genetic research. First, bacteria and their viruses generally have a short generation time. Some viruses increase three-hundredfold in about a half hour; an *Escherichia coli* cell divides every twenty minutes. In contrast, generation time is fourteen days in fruit flies, a year in corn, and twenty years or so in human beings. (*E. coli,* the common intestinal bacterium, was discovered by Theodor Escherich in 1885.)

Another reason bacteria and bacterial viruses are so well-suited for genetic research is because they have much less genetic material than eukaryotes do, and the organization of this material is much simpler. The term *prokaryote* arises from the lack of a true nucleus (*pro* means before and *karyon* means kernel or nucleus); they have no nuclear membranes (see fig. 3.2) and usually have only a single, relatively "naked" chromosome, so they are haploid. Bacteria may, however, contain small, auxiliary circles of DNA, called **plasmids.** Bacterial viruses are even simpler. Although animal and plant viruses, discussed in more detail later in the book (chapters 13 and 16), can be more complicated, the viruses we are interested in studying in this chapter—the bacterial viruses, **bacteriophages,** or just **phages** (Greek: eating)—are exclusively genetic material surrounded by a protein coat.

Bacteriophages are usually classified first by the type of genetic material (nucleic acid) they have (DNA or RNA, single- or double-stranded), then by structural features of their protein surfaces **(capsids)** such as type or symmetry and number of discrete protein subunits **(capsomeres)** in the capsid, and general size. Most bacteriophages are complex, like T2 (fig. 7.1), or made up of a headlike capsule like T2 without the tail appendages, or filamentous. Most contain double-stranded DNA. Bacteriophages are obligate parasites; outside of a host, they are inert molecules. Once their genetic material penetrates a host cell, they can take over the metabolism of that cell and construct multiple copies of themselves. We will discuss details of this and alternative infection

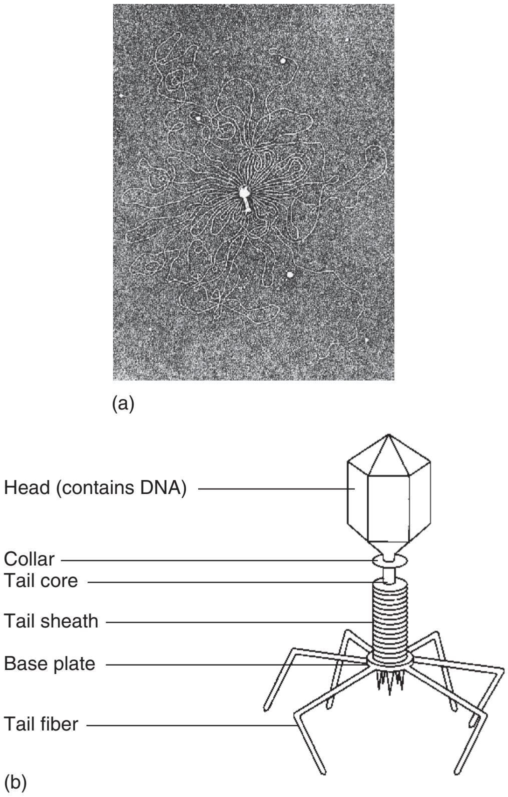


**Figure 7.1** Phage T2 and its chromosome. (*a*) The chromosome, which is about 50 μm long, has burst from the head. (*b*) The intact phage. The phage attaches to a bacterium using its tail fibers and base plate and then injects its genetic material into the host cell. ([*a*] A. K. Kleinschmidt, et al., "Darstellung und Langen messungen des gesamten Deoxyribose-nucleinsaure Inhaltes von T2-Bacteriophagen" *Biochemica et Biophysica Acta,* 61:857–64, 1962. Reproduced by permission of Elsevier Science Publishers.)

pathways later in the chapter. The smallest bacteriophages (e.g., R17) have RNA as their genetic material and contain just three genes, one each for a coat protein, an attachment protein, and an enzyme to replicate their RNA. The larger bacteriophages (T2, T4) have DNA as their genetic material and contain up to 130 genes.

A third reason for the use of bacteria and viruses in genetic study is their ease of handling. A researcher can handle millions of bacteria in a single culture with a minimal amount of work compared with the effort required to grow the same number of eukaryotic organisms such as fruit flies or corn. (Some eukaryotes, such as yeast or *Neurospora,* can, of course, be handled using prokaryotic techniques, as we saw in chapter 6.) Let us look at an expansion of the techniques, introduced in chapter 6, that geneticists use in bacterial and viral studies.

## TECHNIQUES OF CULTIVATION

All organisms need an energy source, a carbon source, nitrogen, sulfur, phosphorus, several metallic ions, and water. Those that require an organic form of carbon are termed **heterotrophs.** Those that can utilize carbon as carbon dioxide are termed **autotrophs.** All bacteria obtain their energy either by photosynthesis or chemical oxidation. Bacteria are usually grown in or on a chemically defined **synthetic medium,** either in liquid in flasks or test tubes, or on petri plates using an agar base to supply rigidity. When one cell is placed on the medium in the plate, it will begin to divide. After incubation, often overnight, a colony, or clone, will exist where previously was only one cell. Overlapping colonies form a confluent growth (fig. 7.2). A culture medium that has only the minimal necessities required by the bacterial species is referred to as *minimal medium* (table 7.1).

Alternatively, bacteria can grow on a medium that supplies, in addition to their minimal requirements, the more complex substances that the bacteria normally synthesize, including amino acids, vitamins, and so on. A medium of this kind allows the growth of strains of bacteria, called **auxotrophs,** that have particular nutritional requirements. (The parent, or wild-type, strain is referred to as a **prototroph.**) For example, a strain that has an enzyme defect in the pathway that produces the amino acid histidine will not grow on a minimal medium because it has no way of obtaining histidine; it is a histidine-requiring auxotroph. If, however, histidine were provided in the medium, the organisms could grow. This type of mutant is called a **conditional-lethal mutant.** The organism would normally die, but under appropriate conditions, such as the addition of histidine, the organism can survive.

This histidine-requiring auxotrophic mutant can grow only on an **enriched** or **complete medium,** whereas the parent prototroph could grow on a minimal medium. Media are often enriched by adding complex mixtures of organic substances such as blood, beef extract, yeast extract, or peptone, a digestion product. Many media, however, are made up of a minimal medium with the addition of only one other substance, such as an amino acid or a vitamin. These are called **selective media;** we will discuss their uses later in the chapter. In addition to minimal, complete, and selective media, other media exist for specific purposes such as aiding in counting colonies, helping maintain cells in a nongrowth phase, and so on.

**Table 7.1** **Minimal Synthetic Medium for Growing *E. coli,* a Heterotroph**

| Component | Quantity |
|---|---|
| $NH_4H_2PO_4$ | 1 g |
| Glucose | 5 g |
| NaCl | 5 g |
| $MgSO_4 \cdot 7H_2O$ | 0.2 g |
| $K_2HPO_4$ | 1 g |
| $H_2O$ | 1,000 ml |

*Source:* Data from M. Rogosa, et al., *Journal of Bacteriology,* 54:13, 1947.

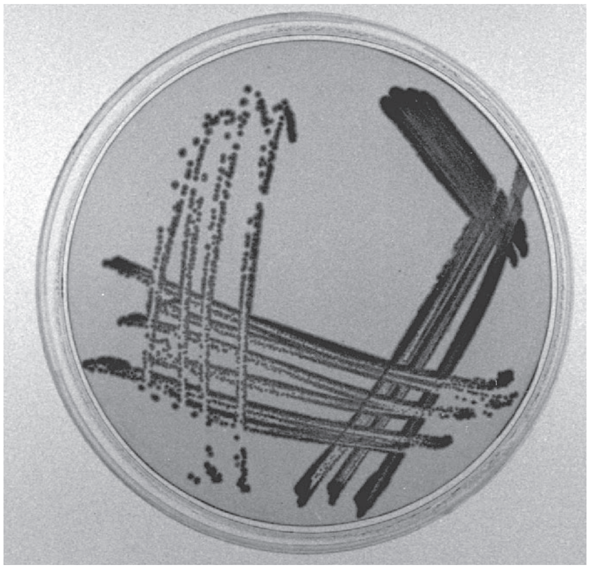


**Figure 7.2** Confluent growth of bacterial colonies on a petri plate. Bacteria were streaked on the petri plate with an inoculation loop—a metal wire with a looped end—covered with bacteria. Streaks began at the upper right and continued around clockwise. With a heavy inoculation on the loop, bacterial growth is confluent. Eventually, only a few bacteria are left; they form single colonies at the upper left. (Photo by Robert Tamarin.)

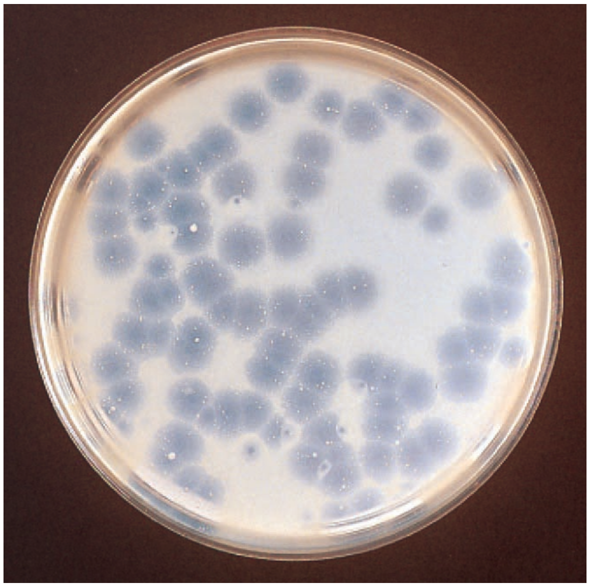


**Figure 7.3** Viral plaques (phage T1) on a bacterial lawn of *E. coli.* (© Bruce Iverson, BSc.)

The experimental cultivation of viruses is somewhat different. Since viruses are obligate parasites, they can grow only in living cells. Thus, for the cultivation of phages, petri plates of appropriate media are inoculated with enough bacteria to form a continuous cover, or **bacterial lawn.** This bacterial culture serves as a medium for the growth of viruses added to the plate. Since the virus attack usually results in rupture, or **lysis,** of the bacterial cell, addition of the virus usually produces clear spots, known as **plaques,** on the petri plates (fig. 7.3). Large quantities of viruses can be grown in flasks of bacteria.

## BACTERIAL PHENOTYPES

Bacterial phenotypes fall into three general classes: colony morphology, nutritional requirements, and drug or infection resistance.

### Colony Morphology

The first of these classes, colony morphology, relates simply to the form, color, and size of the colony that grows from a single cell. A bacterial cell growing on a petri plate in an incubator at 37° C divides as frequently as once every twenty minutes. Each cell gives rise to a colony, or clone, at its original position. In a relatively short amount of time (e.g., overnight), the colonies will consist of enough cells to be seen with the unaided eye. The different morphologies observed among the colonies are usually under genetic control (fig. 7.4).

### Nutritional Requirements

The second basis for classifying bacteria—by their nutritional requirements—reflects the failure of one or more enzymes in the bacteria's biosynthetic pathways. If an auxotroph has a requirement for the amino acid cysteine that the parent strain (prototroph) does not have, then that auxotroph most likely has a nonfunctional enzyme in the pathway for the synthesis of cysteine. Figure 7.5 shows five steps in cysteine synthesis; a different enzyme controls each step. All enzymes are proteins, and the information in one or more genes determines the sequences in the strings of amino acids that make up those proteins (chapter 11). A normal or wild-type allele

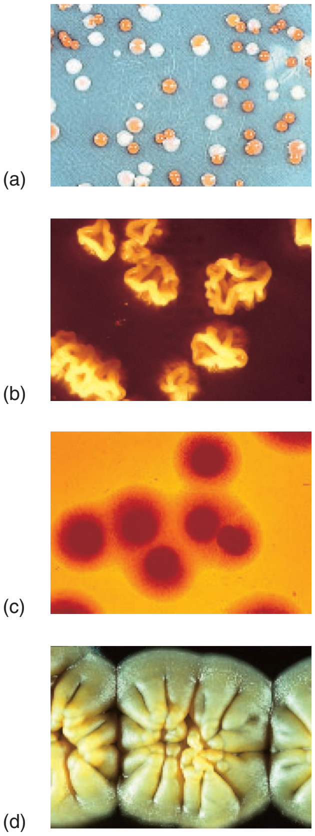


**Figure 7.4** Various bacterial colony forms on agar petri plates. (*a*) Red and white colonies of *Serratia marcescens.* (*b*) Irregular raised folds of *Streptomyces griseus*. (*c*) Round colonies with concentrated centers and diffuse edges of *Mycoplasma.* (*d*) Irregularly folded raised colonies of *Streptomyces antibioticus.* ([*a*] © Dr. E. Buttone/Peter Arnold, Inc., [*b*] © C. Case/Visuals Unlimited, [*c*] © Michael G. Gabridge/Visuals Unlimited, [*d*] © Cabisco/Visuals Unlimited.)

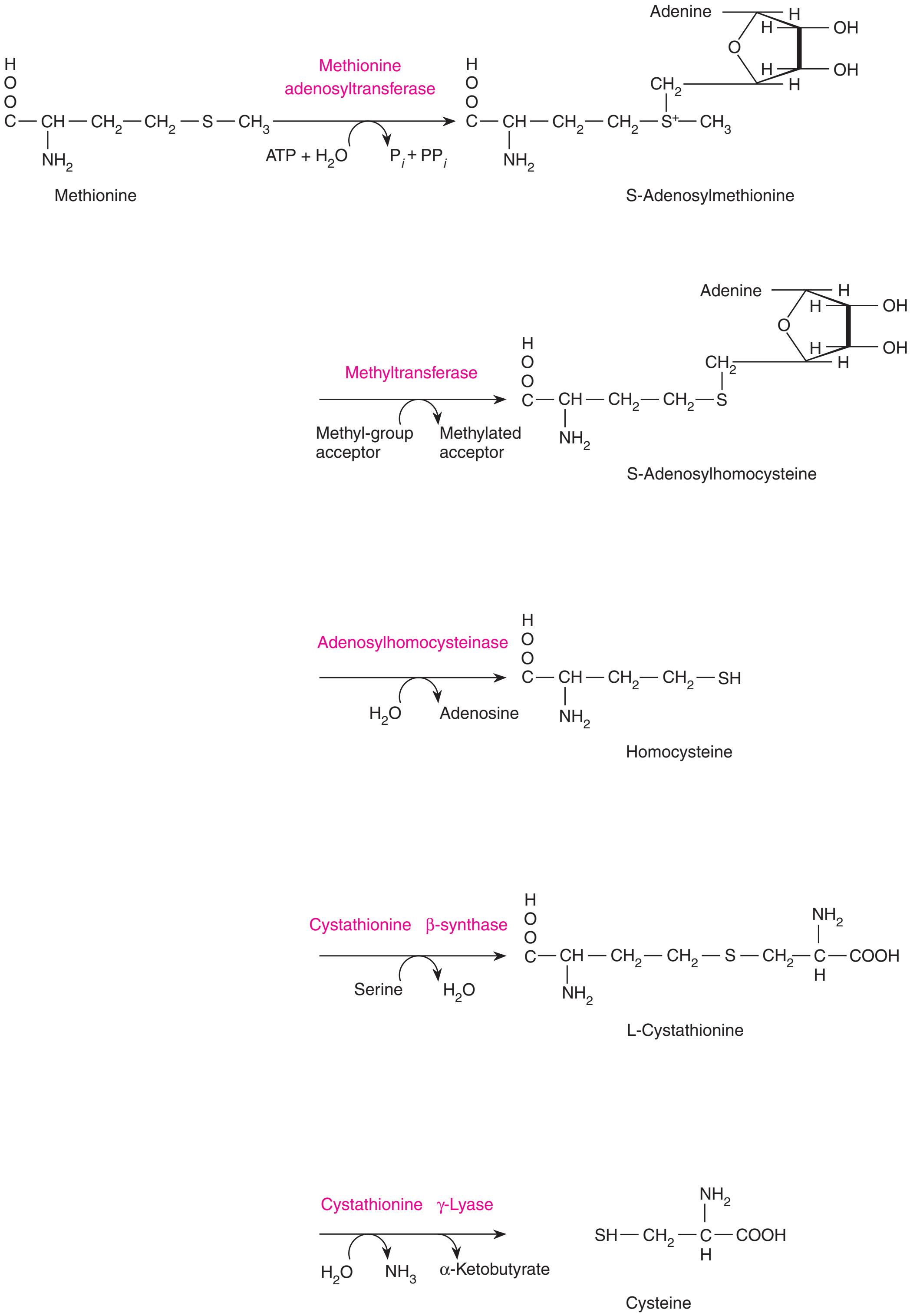
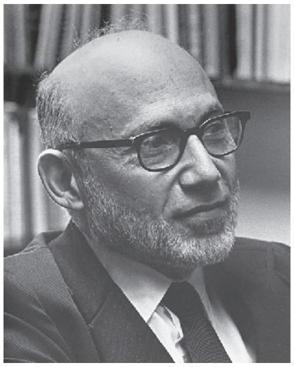

Methionine
Methionine adenosyltransferase
ATP + $H_2O$
$P_i + PP_i$
S-Adenosylmethionine
Adenine
Methyltransferase
Methyl-group acceptor
Methylated acceptor
S-Adenosylhomocysteine
Adenosylhomocysteinase
$H_2O$
Adenosine
Homocysteine
Cystathionine β-synthase
Serine
$H_2O$
L-Cystathionine
Cystathionine γ-Lyase
$H_2O$
$NH_3$
α-Ketobutyrate
Cysteine

**Figure 7.5** Five-step conversion of methionine to cysteine. Each step is controlled by a different enzyme (*red*).



Joshua Lederberg (1925– ).
(Courtesy of Dr. Joshua Lederberg.)

produces a normal, functional enzyme. The alternative allele may produce a nonfunctional enzyme. Recall the one-gene-one-enzyme hypothesis from chapter 2.

A technique known as **replica-plating,** devised by Joshua Lederberg, is a rapid **screening technique** that makes it possible to determine quickly whether a given strain of bacteria is auxotrophic for a particular metabolite. In this technique, a petri plate of complete medium is inoculated with bacteria. The resulting growth will have a certain configuration of colonies. This plate of colonies is pressed onto a piece of sterilized velvet. Then any number of petri plates, each containing a medium that lacks some specific metabolite, can be pressed onto this velvet to pick up inocula in the same pattern as the growth on the original plate (fig. 7.6). If a colony grows on the complete medium but does not grow on a plate with a medium missing a metabolite, the inference is that the colony is made of auxotrophic cells that require the absent metabolite. Samples of this bacterial strain can be obtained from the colony growing on complete medium for further study. The nutritional requirement of this strain is its phenotype. The methionine-requiring auxotroph of figure 7.6 would be designated as $Met^-$ (methionine-minus or Met-minus).

In terms of energy sources, the plus or minus notation has a different meaning. For example, a strain of bacteria that can utilize the sugar galactose as an energy source would be $Gal^+$. If it could not utilize galactose, it would be called $Gal^-$. The latter strain will not grow if galactose is its sole carbon source. It will grow if a sugar other than galactose is present. Note that a $Met^-$ strain needs methionine to grow, whereas a $Gal^-$ strain needs a carbon source *other* than galactose; it cannot use galactose.

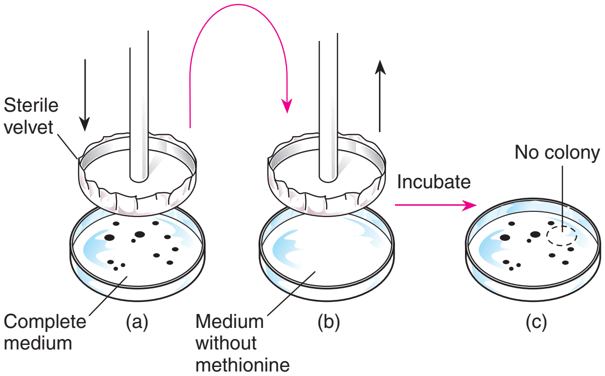


**Figure 7.6** Replica-plating technique. (*a*) A pattern of colonies from a plate of complete medium is transferred (*b*) to a second plate of medium that lacks methionine. (*c*) In the locations where colonies fail to grow on the second plate, we can infer that the original colony was a methionine-requiring auxotroph.

## *Resistance and Sensitivity*

The third common classification of phenotypes in bacteria involves resistance and sensitivity to drugs, phages, and other environmental insults. For example, penicillin, an antibiotic that prevents the final stage of cell-wall construction in bacteria, will kill growing bacterial cells. Nevertheless, we frequently find a number of cells that do grow in the presence of penicillin. These colonies are resistant to the drug, and this resistance is under simple genetic control. The phenotype is penicillin resistant ($Pen^r$) as compared with penicillin sensitive ($Pen^s$), the normal condition, or wild-type. Numerous antibiotics are used in bacterial studies (table 7.2).

**Table 7.2** **Some Antibiotics and Their Antibacterial Mechanisms**

| Antibiotic | Microbial Origin | Mode of Action |
|---|---|---|
| Penicillin G | *Penicillium chrysogenum* | Blocks cell-wall synthesis |
| Tetracycline | *Streptomyces aureofaciens* | Blocks protein synthesis |
| Streptomycin | *Streptomyces griseus* | Interferes with protein synthesis |
| Terramycin | *Streptomyces rimosus* | Blocks protein synthesis |
| Erythromycin | *Streptomyces erythraeus* | Blocks protein synthesis |
| Bacitracin | *Bacillus subtilis* | Blocks cell-wall synthesis |

Drug sensitivity provides another screening technique for isolating nutritional mutations. For example, if we were looking for mutants that lacked the ability to synthesize a particular amino acid (e.g., methionine), we could grow large quantities of bacteria (prototrophs) and then place them on a medium that lacked methionine but contained penicillin. Here, any growing cells would be killed. But methionine auxotrophs would not grow, and, therefore, they would not be killed. The penicillin could then be washed out and the cells reinoculated onto a complete medium. The only colonies that form should be composed of methionine auxotrophs ($Met^-$).

Screening for resistance to phages is similar to screening for drug resistance. When bacteria are placed in a medium containing phages, only those bacteria that are resistant to the phages will grow and produce colonies. They can thus be easily isolated and studied.

## VIRAL PHENOTYPES

Bacteriophage phenotypes fall generally into two categories: plaque morphology and growth characteristics on different bacterial strains. For example, T2, an *E. coli* phage (see fig. 7.1), produces small plaques with fuzzy edges (genotype $r^+$). Rapid-lysis mutants (genotype $r$) produce large, smooth-edged plaques (fig. 7.7). Similarly, T4, another *E. coli* phage, has rapid-lysis mutants that produce large, smooth-edged plaques on *E. coli* B but will not grow at all on *E. coli* K12, a different strain. Here, rapid-lysis mutants illustrate both the colony morphology phenotypes and the growth-restriction phenotypes of phages.

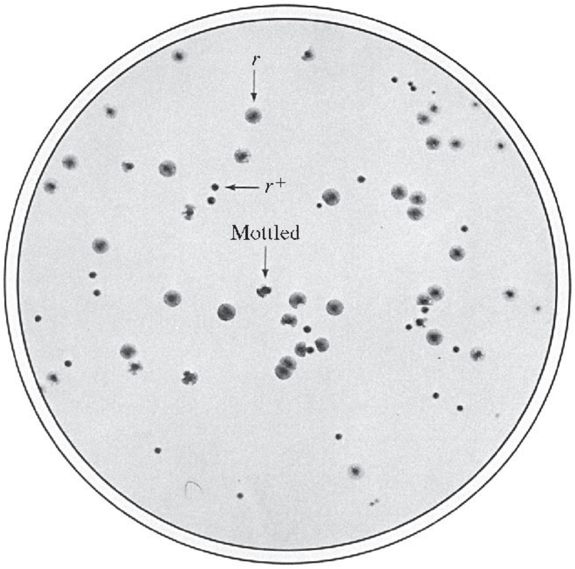


**Figure 7.7** Normal ($r^+$) and rapid-lysis ($r$) mutants of phage T2. Mottled plaques occur when $r$ and $r^+$ phages grow together.
(From *Molecular Biology of Bacterial Viruses* by Gunther S. Stent. © 1963, 1978 by W. H. Freeman and Company. Used with permission.)

## SEXUAL PROCESSES IN BACTERIA AND BACTERIOPHAGES

Although bacteria and viruses are ideal subjects for biochemical analysis, they would not be useful for genetic study if they did not have sexual processes. If we define a sexual process as combining genetic material from two individuals, then the life cycles of bacteria and viruses include sexual processes. Although they do not undergo sexual reproduction by the fusion of haploid gametes, bacteria and viruses do undergo processes that incorporate genetic material from one cell or virus into another cell or virus, forming recombinants. Actually, bacteria have three different methods to gain access to foreign genetic material: **transformation, conjugation,** and **transduction** (fig. 7.8).

Phages can exchange genetic material when a bacterium is infected by more than one virus particle **(virion).** During the process of viral infection, the genetic material of different phages can exchange parts (or recombine; see fig. 7.8). We will examine the exchange processes in bacteria and then in bacteriophages, and then proceed to the use of these methods for mapping bacterial and viral chromosomes. (*Chromosome* refers to the structural entity in the cell or virus made up of the genetic material. In eukaryotes, it is double-stranded DNA complexed with proteins [chapter 15]. Staining of this eukaryotic organelle led to the term *chromosome,* which means "colored body." In prokaryotes, the chromosome is a circle [usually] of double-stranded DNA. In viruses, it is virtually any combination of linear or circular, single- or double-stranded RNA or DNA. Sometimes the term **genophore** is used for the prokaryotic and viral genetic material, limiting the word *chromosome* to the eukaryotic version. We will use the term *chromosome* for the intact genetic material of any organism or virus.)

### Transformation

Transformation was first observed in 1928 by F. Griffith and was examined at the molecular level in 1944 by O. Avery and his colleagues, who used the process to demonstrate that DNA was the genetic material of bacteria. Chapter 9 presents the details of these experiments. In transformation, a cell takes up extraneous DNA found in the environment and incorporates it into its genome

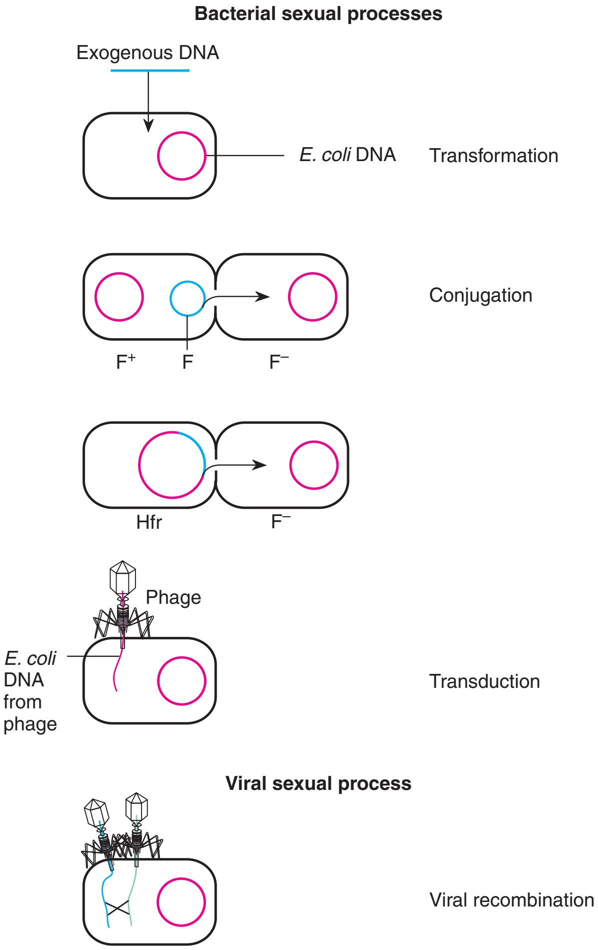


**Figure 7.8** Summary of bacterial and viral sexual processes.

(genetic material) through recombination. Not all bacteria are competent to be transformed, and not all extracellular DNA is competent to transform. To be competent to transform, the extracellular DNA must be double-stranded. To be competent to be transformed, a cell must have the surface protein, **competence factor,** which binds to the extracellular DNA in an energy-requiring reaction. However, bacteria that are not naturally competent can be treated to make them competent, usually by treatment with calcium chloride, which makes them more permeable.

### Mechanisms of Transformation

Under natural conditions, only one of the strands of extracellular DNA is brought into the cell. The single strand brought into the cell can then be incorporated into the host genome by two crossovers (fig. 7.9). (The molecular mechanisms of crossing over are presented in chapter 12.) Note that unlike eukaryotic crossing over, this is not a reciprocal process. The bacterial chromosome incorporates part of the foreign DNA. The remaining single-stranded DNA, originally part of the bacterial chromosome, is degraded by host enzymes called exonucleases; linear DNA is degraded rapidly in prokaryotes.

Transformation is a very efficient method of mapping in some bacteria, especially those that are inefficient in other mechanisms of DNA intake (such as transduction, discussed later in this chapter). For example, a good deal of the mapping of the soil bacterium, *Bacillus subtilis,* has been done through the process of transformation; *E. coli,* however, is inefficient in transformation, so other methods are used to map its chromosome.

### Transformation Mapping

The general idea of transformation mapping is to add DNA from a bacterial strain with known genotype to another strain, also with known genotype, but with different alleles at two or more loci. We then look for incorporation of the donor alleles into the recipient strain of bacteria. The more often alleles from two loci are incorporated together into the host, the closer together these loci must be to each other. Thus, we can use an index of co-occurrence that is in inverse relationship to map distance: the larger the co-occurrence of alleles of two loci, the closer together the loci must be. This is another way of looking at the mapping concepts we discussed in

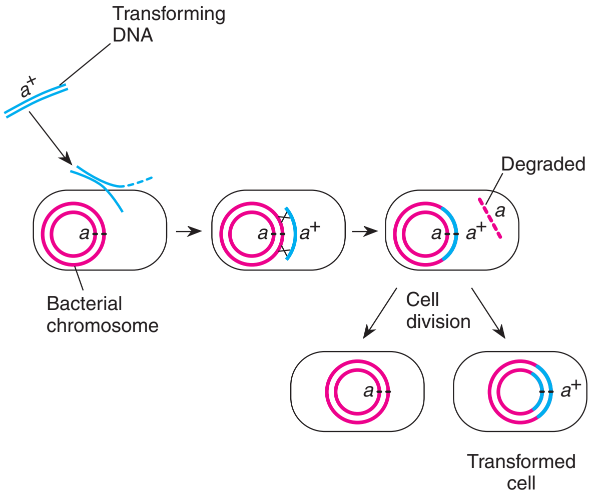


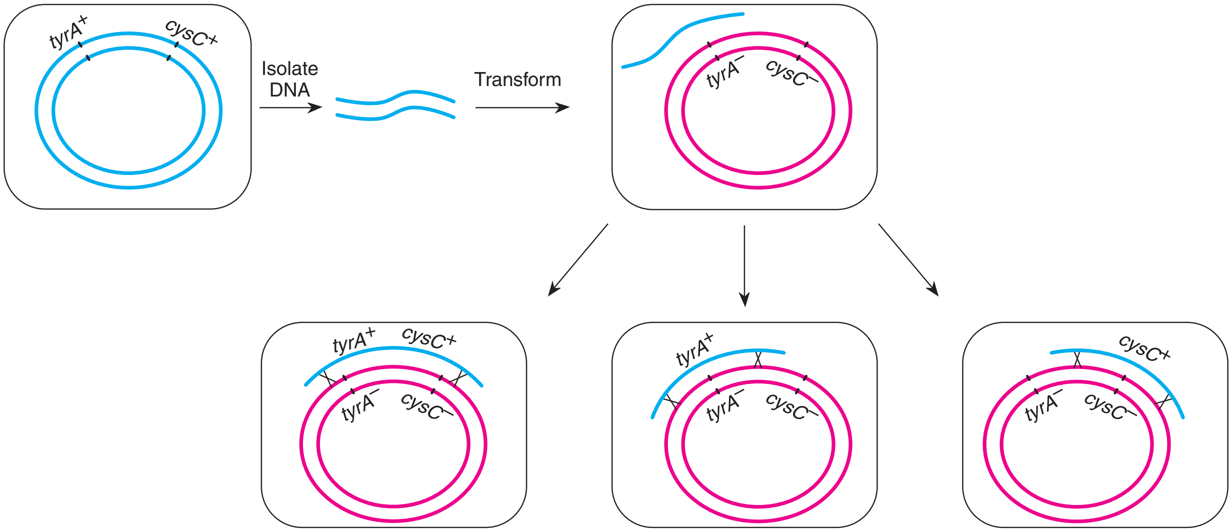
**Figure 7.9** A single strand of transforming DNA (*blue* with $a^+$ allele) enters a bacterial cell (*red* chromosome with *a* allele). Two crossovers bring the foreign DNA into the bacterial chromosome. After DNA replication and cell division, one cell has the *a* allele and the other the $a^+$ allele. The chromosome is drawn as a double circle, symbolizing the double-stranded structure of DNA.

chapter 6, where we discovered that the closer two loci are, the fewer the recombinations between them and thus the higher the co-occurrence.

Now, we also must look at another concept, that is, selecting for recombinant cells. In fruit flies, every offspring of a mated pair represents a sampling of the meiotic tetrad, and thus a part of the total, whether or not recombination took place. Here, however, many cells are present that do not take part in the transformation process. In a bacterial culture, for example, only one cell in a thousand might be transformed. We must thus always be sure when working with bacterial gene transfer that we count only those cells that have taken part in the process. Let us look at an example.

A recipient strain of *B. subtilis* is auxotrophic for the amino acids tyrosine (*tyrA*$^-$) and cysteine (*cysC*$^-$). We are interested in how close these loci are on the bacterial chromosome. We thus isolate DNA from a prototrophic strain of bacteria (*tyrA*$^+$ *cysC*$^+$). We add this donor DNA to the auxotrophic strain and allow time for transformation to take place (fig. 7.10). If the experiment is successful, and the loci are close enough together, then some of the recipient bacteria may incorporate donor DNA that has either both donor alleles or one or the other donor allele. Thus, some of the recipient cells will now have the *tyrA*$^+$ and *cysC*$^+$ alleles, some will have just the donor *tyrA*$^+$ allele, some will have just the donor *cysC*$^+$ allele, and the overwhelming majority will be of the untransformed auxotrophic genotype, *tyrA*$^-$ *cysC*$^-$. We thus need to count the transformed cells.

We do this by removing any extraneous transforming DNA and then pouring the cells out onto a complete medium so that all cells can grow. These cells are then replica-plated onto three plates—a minimal medium plate, a minimal medium plus tyrosine plate, and a minimal medium plus cysteine plate—and allowed to grow overnight in an incubator at 37° C. We then count colonies (fig. 7.11). Those growing on minimal medium are of genotype *tyrA*$^+$ *cysC*$^+$; those growing on minimal medium with tyrosine but not growing on minimal medium are *tyrA*$^-$ *cysC*$^+$; and those growing on minimal medium with cysteine but not growing on minimal medium are *tyrA*$^+$ *cysC*$^-$. The overwhelming majority will grow on complete medium, but not on minimal medium or minimal media with just tyrosine or cysteine added. This majority is made up of the nontransformants, that is, auxotrophs that were not involved in a transformation event—they took up no foreign DNA.



| Growth | *tyrA*$^+$ *cysC*$^+$ | *tyrA*$^+$ *cysC*$^-$ | *tyrA*$^-$ *cysC*$^+$ |
|---|---|---|---|
| Complete medium | + | + | + |
| Minimal medium | + | – | – |
| Minimal + tyrosine | + | – | + |
| Minimal + cysteine | + | + | – |

**Figure 7.10** Transformation experiment with *B. subtilis.* A *tyrA*$^-$ *cysC*$^-$ strain is transformed with DNA from a *tyrA*$^+$ *cysC*$^+$ strain. Nontransformants as well as three types of transformants (two single and one double) result. Genotypes are determined by growth characteristics on four different types of petri plates (see fig. 7.11).

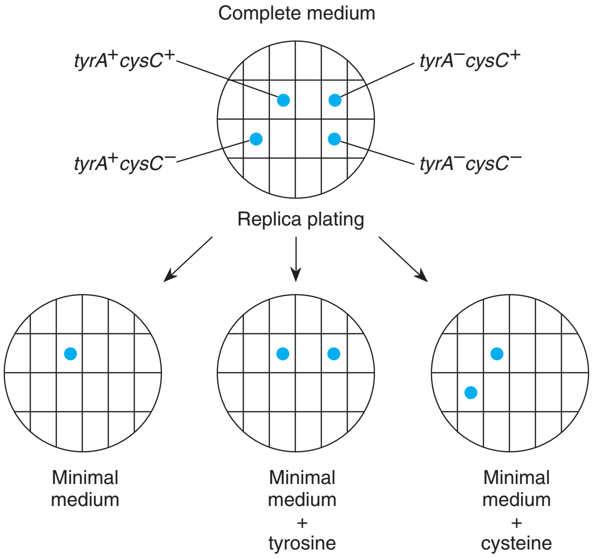


**Figure 7.11** Four patterns of growth on different media reveal the genotypes of transformed and untransformed cells. Only four colonies are shown, and a grid is added for ease of identification. After transformation (see fig. 7.10), cells are plated on complete medium and then replica-plated onto minimal medium with either tyrosine or cysteine added.

As a control against reversion, the normal mutation of *tyrA*$^-$ to *tyrA*$^+$ or *cysC*$^-$ to *cysC*$^+$, we grow several plates of auxotrophs in minimal medium and minimal medium with tyrosine or cysteine added. These are auxotrophs that were not exposed to prototrophic donor DNA. We then count the number of natural revertants and correct our experimental numbers by the natural reversion rate. Thus, we are sure that what we measure is the actual transformation rate rather than just a mutation rate that we mistake for transformation. This control should *always* be carried out.

From the experiment (see figs. 7.10 and 7.11), we count twelve double transformants (*tyrA*$^+$ *cysC*$^+$), thirty-one *tyrA*$^+$ *cysC*$^-$, and twenty-seven *tyrA*$^-$ *cysC*$^+$. From these data, we calculate the co-occurrence, or cotransfer index, ($r$) as

$$r = \frac{\text{number of double transformants}}{\text{number of double transformants} + \text{number of single transformants}}$$

From our data

$$r = 12/(12 + 31 + 27) = 0.17.$$

This is a relative number indicating the co-occurrence of the two loci and thus their relative distance apart on the bacterial chromosome. Remember that as this number increases for different pairs of loci, the loci are closer and closer together.

By systematically examining many loci, we can establish their relative order. For example, if locus *A* is closely linked to locus *B* and *B* to *C,* we can establish the order *A B C.* It is not possible by this method to determine exact order for very closely linked genes. For this information we need to rely upon transduction, which we will consider shortly. However, transformation has allowed us to determine that the map of *B. subtilis* is circular, a phenomenon found in all prokaryotes and many phages. (The *E. coli* map is shown later.)

## Conjugation

In 1946, Joshua Lederberg and Edward L. Tatum (later to be Nobel laureates) discovered that *E. coli* cells can exchange genetic material through the process of conjugation. They mixed two auxotrophic strains of *E. coli.* One strain required methionine and biotin (Met$^-$ Bio$^-$), and the other required threonine and leucine (Thr$^-$ Leu$^-$). This cross is shown in figure 7.12. Remember that if a strain is Met$^-$ Bio$^-$, it is, without saying, wild-type for all other loci. Thus, a cell with the Met$^-$ Bio$^-$ phenotype actually has the genotype of *met*$^-$ *bio*$^-$ *thr*$^+$ *leu*$^+$. Similarly, the Thr$^-$ Leu$^-$ strain is actually *met*$^+$ *bio*$^+$ *thr*$^-$ *leu*$^-$. (Note that symbols such as "Thr$^-$" represent phenotypes; symbols such as "*thr*$^-$" represent genotypes.)

Lederberg and Tatum used multiple auxotrophs in order to rule out spontaneous reversion (mutation). About one in $10^6$ Met$^-$ cells will spontaneously become prototrophic (Met$^+$) every generation. However, with multiple auxotrophs, the probability that several loci will simultaneously and spontaneously revert (e.g., *met*$^-$ → *met*$^+$) becomes vanishingly small. (In fact, the control plates in the experiment, illustrated in fig. 7.12, showed no growth for parental double mutants.) After mixing the strains, Lederberg and Tatum found that about one cell in $10^7$ was prototrophic (*met*$^+$ *bio*$^+$ *thr*$^+$ *leu*$^+$).

To rule out transformation, one strain was put in each arm of a U-tube with a sintered glass filter at the bottom. (fig. 7.13). The liquid and large molecules, including DNA, were mixed by alternate application of pressure and suction to one arm of the tube; whole cells did not pass through the filter. The result was that the fluids surrounding the cells, as well as any large molecules (e.g., DNA), could be freely mixed while the cells were kept separate. After cell growth stopped in the two arms (in complete medium), the contents were plated out on minimal medium. There were no prototrophs in either arm. Therefore, cell-to-cell contact was required for the genetic material of the two cells to recombine.

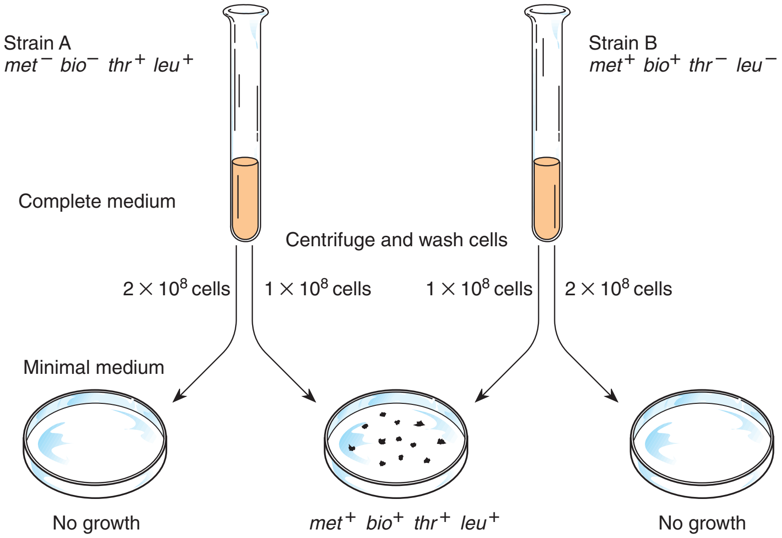


**Figure 7.12** Lederberg and Tatum's cross showing that *E. coli* undergoes genetic recombination.

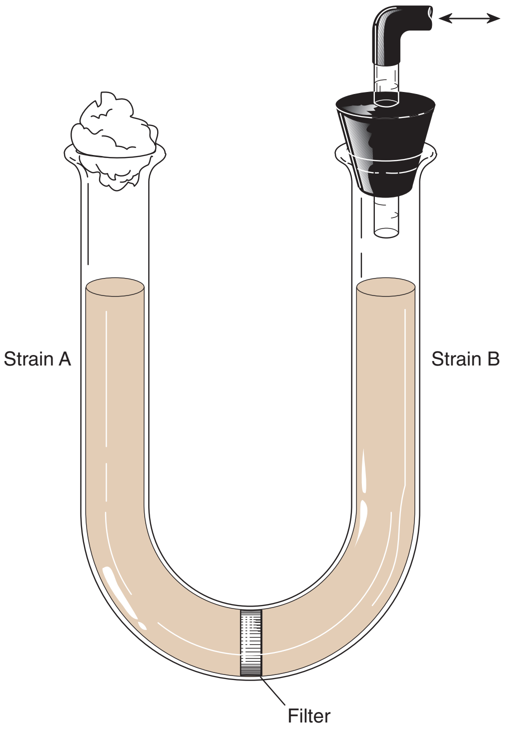


**Figure 7.13** The U-tube experiment. Alternating suction and pressure force liquid and macromolecules back and forth across the filter.

### *F Factor*

In bacteria, conjugation is a one-way transfer, with one strain acting as donor and the other as recipient. Sometimes donor cells, if stored for a long time, lose the ability to be donors, but they can regain the ability if they are mated with other donor strains. This discovery led to the hypothesis that a **fertility factor, F,** made any strain that carried it a male (donor) strain, termed $F^+$. The strain that did not have the F factor, referred to as a female or $F^-$ strain, served as a recipient for genetic material during conjugation. Research supports this hypothesis.

The F factor is a *plasmid,* a term originally coined by Lederberg to refer to independent, self-replicating genetic particles. Plasmids are usually circles of double-stranded DNA. (Plasmids are at the heart of recombinant DNA technology, which is discussed in detail in chapter 13.) They are auxiliary circles of DNA that many bacteria carry. They are usually much smaller than the bacterial chromosome.

Researchers found that the transfer of the F factor occurred far more frequently than the transfer of other genes from the donor. That is, during conjugation, about one recombinant occurred in $10^7$ cells, whereas transfer of the F factor occurred at a rate of about one conversion of $F^-$ to $F^+$ in every five conjugations. An *E. coli* strain was then discovered that transferred its genetic material at a rate about one thousand times that of the normal $F^+$ strain. This strain was called **Hfr,** for *h*igh *f*requency of *r*ecombination. Several other phenomena occurred with this high rate of transfer. First, the ability to transfer the F factor itself dropped to almost zero in this strain. Second, not all loci were transferred at the same rate. Some loci were transferred much more frequently than others.

*Escherichia coli* cells are normally coated with hairlike **pili (fimbriae).** $F^+$ and Hfr cells have one to three additional pili (singular: pilus) called **F-pili,** or sex pili. During conjugation, these sex pili form a connecting bridge between the $F^+$ (or Hfr) and $F^-$ cells (fig. 7.14). Once a connection is made, the sex pilus then contracts to bring the two cells into contact. DNA transfer takes place through a nick in either the plasmid (in $F^+$ cells) or the bacterial chromosome (in Hfr cells). A single strand of the DNA double-stranded donor DNA then passes from the $F^+$ or Hfr cell to the $F^-$ cell across the cell membranes. DNA replication in both the donor and recipient cells reestablishes double-stranded DNA in both. The F factor itself has the genes for sex-pilus formation and DNA transfer to a conjugating $F^-$ cell. At least twenty-two genes are involved in the transfer process, including genes for the pilus protein, nicking the DNA, and regulation of the process.

In the transfer process of conjugation, the donor cell does not lose its F factor or its chromosome because only a single strand of the DNA double helix is transferred; the remaining single strand is quickly replicated. (The process of DNA replication is described in chapter 9.) For a short while, the $F^-$ cell that has conjugated with an Hfr cell has two copies of whatever chromosomal loci were transferred: one copy of its own and one transferred in. With these two copies, the cell is a partial diploid, or a **merozygote.** The new foreign DNA **(exogenote)** can be incorporated into the host chromosome **(endogenote)** by an even number of breakages and reunions between the two, just as in transformation. The unincorporated linear DNA is soon degraded by enzymes. The conjugation process is diagrammed in figure 7.15.

## Interrupted Mating

To demonstrate that the transfer of genetic material from the donor to the recipient cell during conjugation is a linear event, F. Jacob and E. Wollman devised the technique of **interrupted mating.** In this technique, $F^-$ and Hfr strains were mixed together in a food blender.

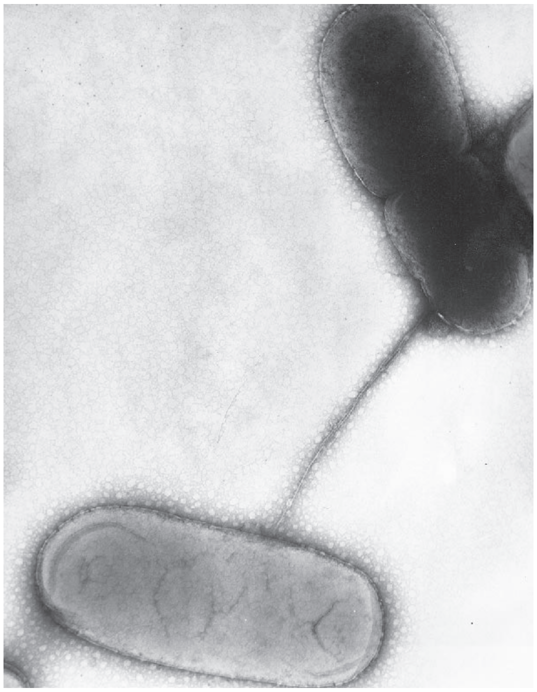


**Figure 7.14** Electron micrograph of conjugation between an $F^+$ (*upper right*) and an $F^-$ (*lower left*) cell with the F-pilus between them. Magnification 3,700×. (Courtesy of Wayne Rosenkrans and Dr. Sonia Guterman.)

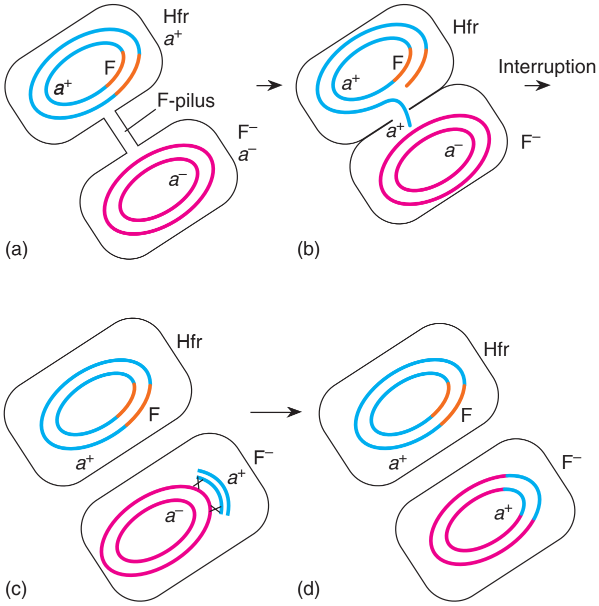


**Figure 7.15** Bacterial conjugation. (*a*) The F-pilus draws an Hfr and an $F^-$ cell close together. (*b*) The Hfr chromosome then begins to pass into the $F^-$ cell, beginning at the F region of the Hfr chromosome but in the direction away from the F factor. Only a single strand passes into the $F^-$ cell; this strand and the single strand remaining in the Hfr cell are replicated. After the process is interrupted (*c*), two crossovers bring the $a^+$ allele into the $F^-$ $a^-$ chromosome (*d*).

Elie Wollman (1917– ). (Courtesy of Dr. Elie Wollman and the Pasteur Institute.)

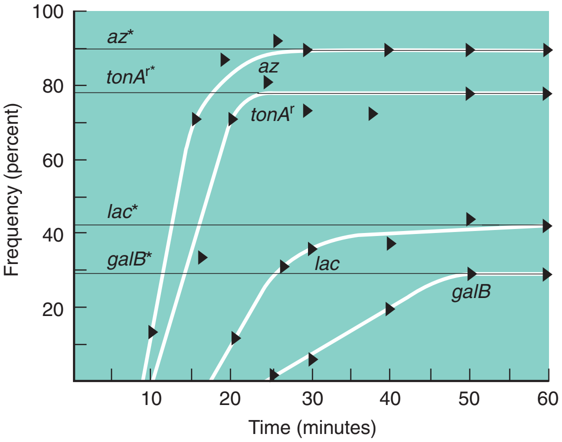
After waiting a specific amount of time, Jacob and Wollman turned the blender on. The spinning motion separated conjugating cells and thereby interrupted their mating. Then the researchers tested the F⁻ cells for various alleles originally in the Hfr cell. In an experiment like this, the Hfr strain is usually sensitive to an antibiotic such as streptomycin. After conjugation is interrupted, the cells are plated onto a medium containing the antibiotic, which kills all the Hfr cells. Then the genotypes of the F⁻ cells can be determined by replica-plating without fear of contamination by Hfr cells.

The mating outlined in table 7.3 was carried out. In the food blender, an Hfr strain sensitive to streptomycin (*str*$^s$) but resistant to azide (*azi*$^r$), resistant to phage T1 (*tonA*$^r$), and prototrophic for the amino acid leucine (*leu*$^+$) and the sugars galactose (*galB*$^+$) and lactose (*lac*$^+$) was added to an F⁻ strain that was resistant to streptomycin (*str*$^r$), sensitive to azide (*azi*$^s$), sensitive to T1 (*tonA*$^s$), and auxotrophic for leucine, galactose, and lactose (*leu*$^-$, *galB*$^-$, and *lac*$^-$). After a specific number of minutes (ranging from zero to sixty), the food blender was turned on. To kill all the Hfr cells, the cell suspension was plated on a medium containing streptomycin. The remaining cells were then plated on medium without leucine. The only colonies that resulted were F⁻ recombinants. They must have received the *leu*$^+$ allele from the Hfr in order to grow on a medium lacking leucine. Hence, all colonies had been selected to be F⁻ recombinants. By replica-plating onto specific media, investigators were able to determine the *azi*, *tonA*, *lac*, and *galB* alleles and the percentage of recombinant colonies that had the original Hfr allele (*leu*$^+$). (Note that by trial and error, it was determined that leucine should be the locus to use to select for recombinants. As we will see, the leucine locus entered first.)

Figure 7.16 shows that as time of mating increases, two things happen. First, new alleles enter the F⁻ cells from the Hfr cells. The *tonA*$^r$ allele first appears among recombinants after about ten minutes of mating, whereas *galB*$^+$ first enters the F⁻ cells after about twenty-five minutes. This suggests a sequential entry of loci into the F⁻ cells from the Hfr (fig. 7.17). Second, as time proceeds, the percentage of recombinants with a given allele from the Hfr increases. At ten minutes, *tonA*$^r$ is first found among recombinants. After fifteen minutes, about 40% of recombinants have the *tonA*$^r$ allele from the Hfr; and after about twenty-five minutes, about 80% of the recombinants have the *tonA*$^r$ allele. This limiting percent-

**Table 7.3** **Genotypes of Hfr and F⁻ Cells Used in an Interrupted Mating Experiment**

| Hfr | F⁻ |
|---|---|
| *str*$^s$ | *str*$^r$ |
| *azi*$^r$ | *azi*$^s$ |
| *tonA*$^r$ | *tonA*$^s$ |
| *leu*$^+$ | *leu*$^-$ |
| *galB*$^+$ | *galB*$^-$ |
| *lac*$^+$ | *lac*$^-$ |



*Limiting percentage for *az*, *tonA*$^r$, *lac*, and *galB* loci.

**Figure 7.16** Frequency of Hfr genetic characters among recombinants after interrupted mating. As time proceeds, new alleles appear and then increase in frequency. Interruption of the mating limits the frequency of successful passage. (From F. Jacobs and E. L. Wollman, *Sexuality and the Genetics of Bacteria,* Academic Press, 1961.)

age does not increase with additional time. The limiting percentage is lower for loci that enter later, a fact explained by the assumption that even without the food blender, mating is usually interrupted before completion by normal agitation alone.

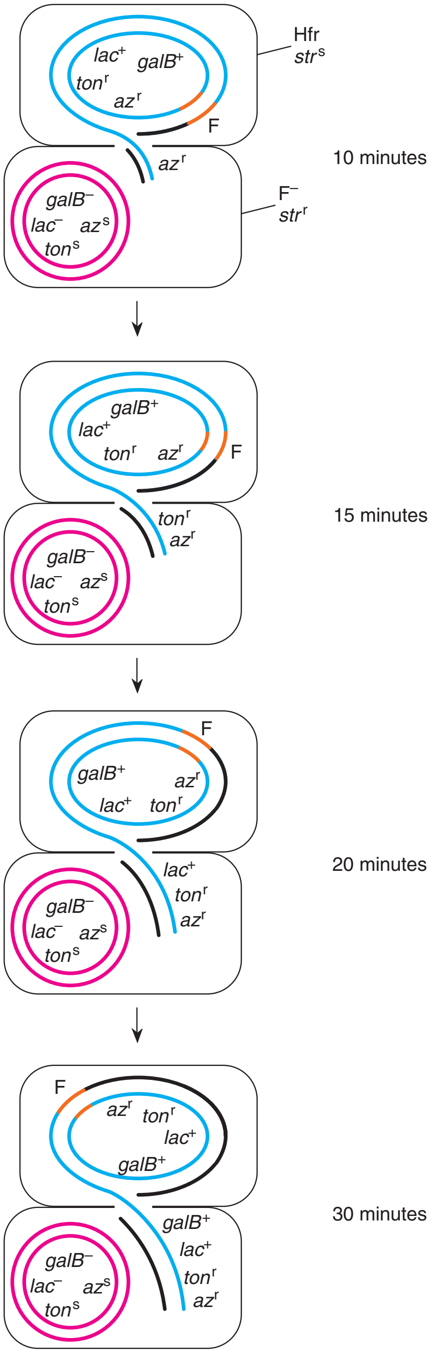


**Figure 7.17** Conjugation in *E. coli.* Hfr chromosome is *blue;* F$^-$ chromosome is *red;* and new DNA replication is *black.* As time proceeds, alleles from the Hfr enter the F$^-$ cell in an orderly, sequential fashion. After the cells separate, two crossovers can bring Hfr alleles into the F$^-$ chromosome. The F factor (*orange*) is the last part of the Hfr chromosome to enter the F$^-$ cell.

## *Mapping and Conjugation*

Jacob and Wollman, working with several different Hfr strains, collected data that indicated that the bacterial chromosome was circular. The strains were of independent origin, and the results were quite striking (table 7.4).

If we ponder this table for a short while, one fact becomes obvious: The relative order of the loci is always the same. What differs is the point of origin and the direction of the transfer. Jacob and Wollman proposed that normally the F factor is an independent circular DNA entity in the F$^+$ cell, and that during conjugation only the F factor is passed to the F$^-$ cell. Since it is a small fragment of DNA, it can be passed entirely in a high proportion of conjugations before the cells separate. Every once in a while, however, the F factor becomes integrated into the chromosome of the host, which then becomes an Hfr cell. The point of integration can be different in different strains. However, once the F factor is integrated, it determines the initiation point of transfer for the *E. coli* chromosome, as well as the direction of transfer.

The F factor is the last part of the *E. coli* chromosome to be passed from the Hfr cell. This explains why an Hfr, in contrast to an F$^+$, rarely passes the F factor itself. In the original work of Lederberg and Tatum, the one recombinant in $10^7$ cells most likely came from a conjugation between an F$^-$ cell and an Hfr that had formed spontaneously from an F$^+$ cell. Integration of the F factor is diagrammed in figure 7.18. The F factor can also reverse this process and loop out of the *E. coli* chromosome. (Sometimes the F factor loops out incorrectly, as in figure 7.19, forming an F′ [F-prime] factor. The passage of this F′ factor to an F$^-$ cell is called **F-duction** or **sexduction.** Not really useful in mapping, the process has proved exceptionally useful in studies of gene expression because of the formation of stable merozygotes, which we will examine in chapter 14.)

We could now diagram the *E. coli* chromosome and show the map location of all known loci. The map units would be in minutes, obtained by interrupted mating. However, at this point, the map would not be complete. Interrupted mating is most accurate in giving the relative position of loci that are not very close to each other. With this method alone, a great deal of ambiguity would arise as to the specific order of very close genes on the chromosome. The remaining sexual process in bacteria, transduction, provides the details that interrupted mating or transformation don't explain.

**Table 7.4** Gene Order of Various Hfr Strains Determined by Means of Interrupted Mating

| Types of Hfr | Order of Transfer of Genetic Characters* | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HfrH | 0 | T | L | Az | $T_1$ | Pro | Lac | Ad | Gal | Try | H | S-G | Sm | Mal | Xyl | Mtl | Isol | M | $B_1$ |
| 1 | 0 | L | T | $B_1$ | M | Isol | Mtl | Xyl | Mal | Sm | S-G | H | Try | Gal | Ad | Lac | Pro | $T_1$ | Az |
| 2 | 0 | Pro | $T_1$ | Az | L | T | $B_1$ | M | Isol | Mtl | Xyl | Mal | Sm | S-G | H | Try | Gal | Ad | Lac |
| 3 | 0 | Ad | Lac | Pro | $T_1$ | Az | L | T | $B_1$ | M | Isol | Mtl | Xyl | Mal | Sm | S-G | H | Try | Gal |
| 4 | 0 | $B_1$ | M | Isol | Mtl | Xyl | Mal | Sm | S-G | H | Try | Gal | Ad | Lac | Pro | $T_1$ | Az | L | T |
| 5 | 0 | M | $B_1$ | T | L | Az | $T_1$ | Pro | Lac | Ad | Gal | Try | H | S-G | Sm | Mal | Xyl | Mtl | Isol |
| 6 | 0 | Isol | M | $B_1$ | T | L | Az | $T_1$ | Pro | Lac | Ad | Gal | Try | H | S-G | Sm | Mal | Xyl | Mtl |
| 7 | 0 | $T_1$ | Az | L | T | $B_1$ | M | Isol | Mtl | Xyl | Mal | Sm | S-G | H | Try | Gal | Ad | Lac | Pro |
| AB311 | 0 | H | Try | Gal | Ad | Lac | Pro | $T_1$ | Az | L | T | $B_1$ | M | Isol | Mtl | Xyl | Mal | Sm | S-G |
| AB312 | 0 | Sm | Mal | Xyl | Mtl | Isol | M | $B_1$ | T | L | Az | $T_1$ | Pro | Lac | Ad | Gal | Try | H | S-G |
| AB313 | 0 | Mtl | Xyl | Mal | Sm | S-G | H | Try | Gal | Ad | Lac | Pro | $T_1$ | Az | L | T | $B_1$ | M | Isol |

*Source:* From F. Jacobs and E. L. Wollman, *Sexuality and the Genetics of Bacteria,* Academic Press, 1961.

* The 0 refers to the origin of transfer.

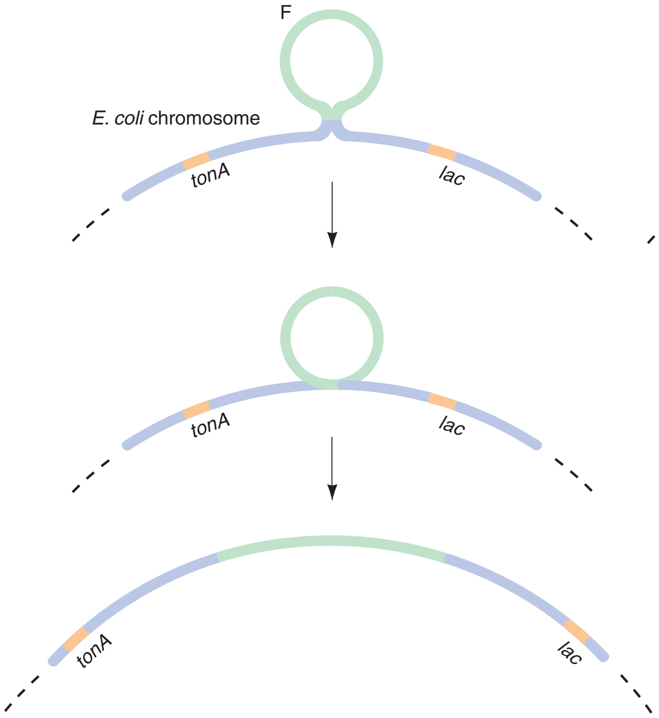


**Figure 7.18** Integration of the F factor by a single crossover. After a simultaneous breakage in both the F factor and the *E. coli* chromosome, the two broken circles reunite to make one large circle, the Hfr chromosome. In this case, the integration occurs between the *tonA* and *lac* loci.

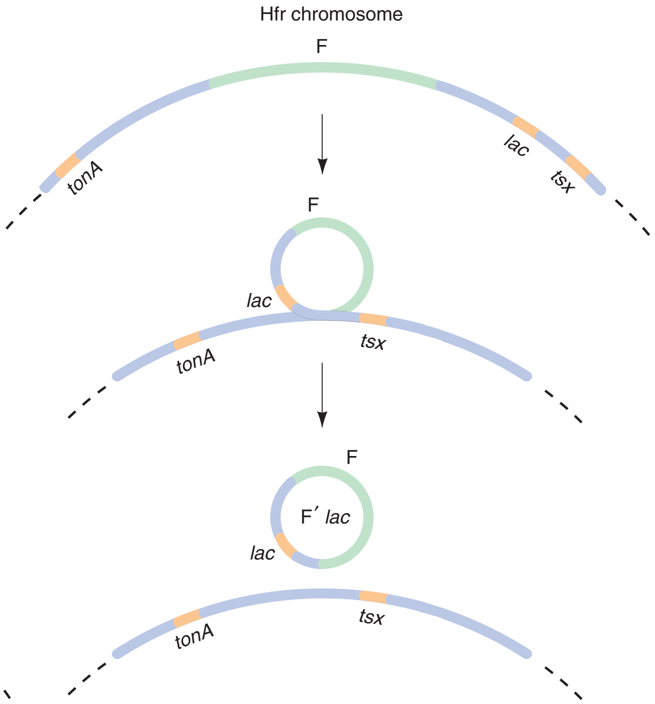


**Figure 7.19** Occasionally, the F factor loops out imprecisely, taking part of the cell's genome in the loop. The circular F factor is freed by a single recombination (crossover) at the loop point.

## LIFE CYCLES OF BACTERIOPHAGES

Phages are obligate intracellular parasites. Phage genetic material enters the bacterial cell after the phage has adsorbed to the cell surface. Once inside, the viral genetic material takes over the metabolism of the host cell. During the infection process, the cell's genetic material is destroyed, while the viral genetic material is replicated many times. The viral genetic material then controls the mass production of various protein components of the virus. New virus particles are assembled within the host cell, which bursts open (is lysed), releasing a **lysate** of hundreds of viral particles to infect other bacteria. This life cycle appears in figure 7.20.

### Recombination

Much genetic work on phages has been done with a group of seven *E. coli* phages called the T series (T-odd: T1, T3, T5, T7; T-even: T2, T4, and T6) and several others, including phage λ (lambda; fig. 7.21). Figure 7.1 diagrammed the complex structure of T2. Phages can undergo recombination processes when a cell is infected with two genetically distinct virions. Hence, the phage genome can be mapped by recombination. As an example, consider the host-range and rapid-lysis loci. Rapid-lysis mutants (*r*) of the T-even phages produce large, sharp-edged plaques. The wild-type produces a smaller, more fuzzy-edged plaque (see fig. 7.7).

Alternative alleles are known also for host-range loci, phage loci that determine the strains of bacteria the phage can infect. For example, T2 can infect *E. coli* cells. These phages can be designated as T2h$^+$ for the normal host range. The *E. coli* is then called Tto$^s$, referring to their sensitivity to the T2 phage. In the course of evolution, an *E. coli* mutant arose that is resistant to the normal phage. This mutant strain is named Tto$^r$ for T2 resistance. In the further course of evolution, the phages have produced mutant forms that can grow on the Tto$^r$ strain of *E. coli.* These phage mutants are designated as T2h for host-range mutant. Remember, *host-range* signifies a mutation in the phage genome, whereas *phage resistance* indicates a mutation in the bacterial genome.

In 1945, Max Delbrück (a 1969 Nobel laureate) developed mixed indicators, which can be used to demonstrate four phage phenotypes on the same petri plate (fig. 7.22). A bacterial lawn of mixed Tto$^r$ and Tto$^s$ is grown. On this lawn, the rapid-lysis phage mutants (*r*) produce large plaques, whereas the wild-type ($r^+$) produce smaller plaques. Phages with host-range mutation (*h*) lyse both Tto$^r$ and Tto$^s$ bacteria. They produce the plaques that are clear (but appear dark) in figure 7.22. Since phages with the wild-type host-range allele ($h^+$) can only infect the Tto$^s$ bacteria, they produce turbid plaques. The Tto$^r$ bacteria growing within these plaques (which appear light-colored in fig. 7.22) produce the turbidity.

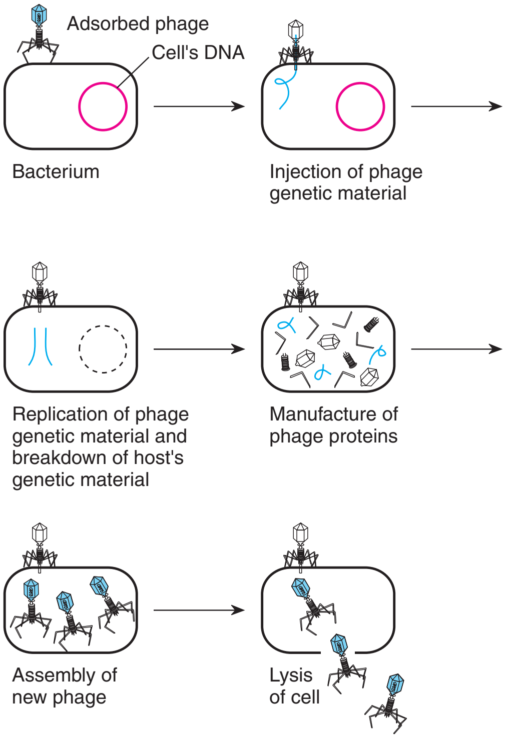

Adsorbed phage
Cell's DNA
Bacterium
Injection of phage genetic material
Replication of phage genetic material and breakdown of host's genetic material
Manufacture of phage proteins
Assembly of new phage
Lysis of cell

**Figure 7.20** The viral life cycle, using T4 infection of *E. coli* as an example.

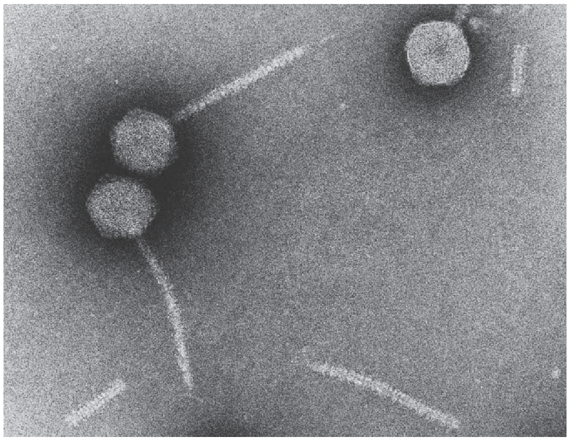


**Figure 7.21** Phage λ. Magnification 167,300×. Note that phage λ lacks the tail fibers and base plate of phage T2 (see fig. 7.1). (Courtesy of Robley Williams.)

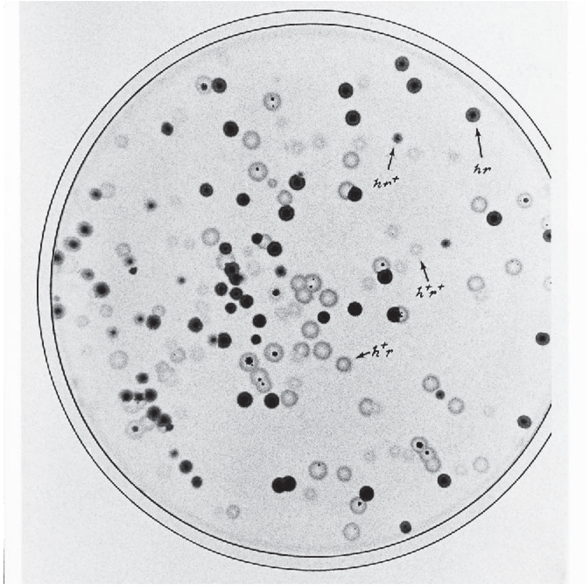


**Figure 7.22** Four types of plaques produced by mixed phage T2 on a mixed lawn of *E. coli.* (From *Molecular Biology of Bacterial Viruses* by Gunther S. Stent, © 1963, 1978 by W. H. Freeman & Company. Used with permission.)

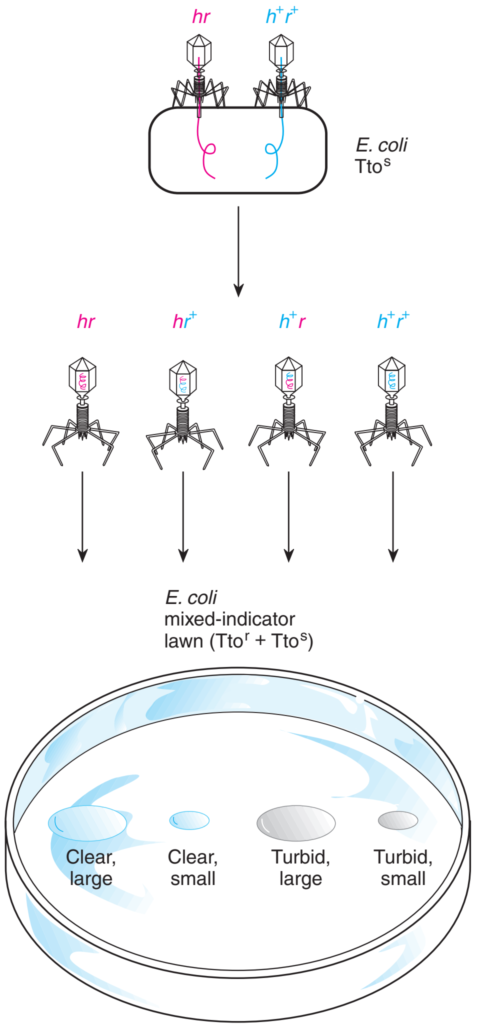


**Figure 7.23** Crossing *hr* and $h^+r^+$ phage. Enough of both types are added to sensitive bacterial cells (Tto$^s$) to ensure multiple infections. The lysate, consisting of four genotypes, is grown on a mixed-indicator bacterial lawn (Tto$^s$ and Tto$^r$). Plaques of four types appear (see fig. 7.22), indicating the genotypes of the parental and recombinant phages.

From the wild stock of phages, we can isolate host-range mutants by looking for plaques on a Tto$^r$ bacterial lawn. Only *h* mutants will grow. These phages can then be tested for the *r* phenotype and the double mutants isolated. Once the two strains (double mutant and wild-type) are available, they can be added in large numbers to sensitive bacteria (fig. 7.23). Large numbers of phages are used to ensure that each bacterium is infected by at least one of each phage type, creating the possibility of recombination within the host bacterium. After a round of phage multiplication, the phages are isolated and plated out on Delbrück's mixed-indicator stock. From this growth, the phenotype (and, hence, genotype) of each phage can be recorded. The percentage of recombinants can be read directly from the plate. For example, on a given petri plate (e.g., fig. 7.22) there might be

| | | | |
|---|---|---|---|
| $hr$ | 46 | $h^+r^+$ | 52 |
| $h^+r$ | 34 | $hr^+$ | 26 |

The first two, *hr* and $h^+r^+$, are the original, or parental, phage genotypes. The second two categories result from recombination between the *h* and *r* loci on the phage chromosome. A single crossover in this region produces the recombinants. Note that with phage recombination, parental phages are counted, since every opportunity was provided for recombination within each bacterium. Thus, every progeny phage arises from a situation in which recombination could have taken place. The proportion of recombinants is

$$(34 + 26)/(46 + 52 + 34 + 26)$$
$$= 60/158 = 0.38 \text{ or } 38\% \text{ or } 38 \text{ map units}$$

This percentage recombination is the map distance, which (as in eukaryotes) is a relative index of distance between loci: The greater the physical distance, the greater the amount of recombination, and thus the larger the map distance. One map unit (1 centimorgan) is equal to 1% recombinant offspring.

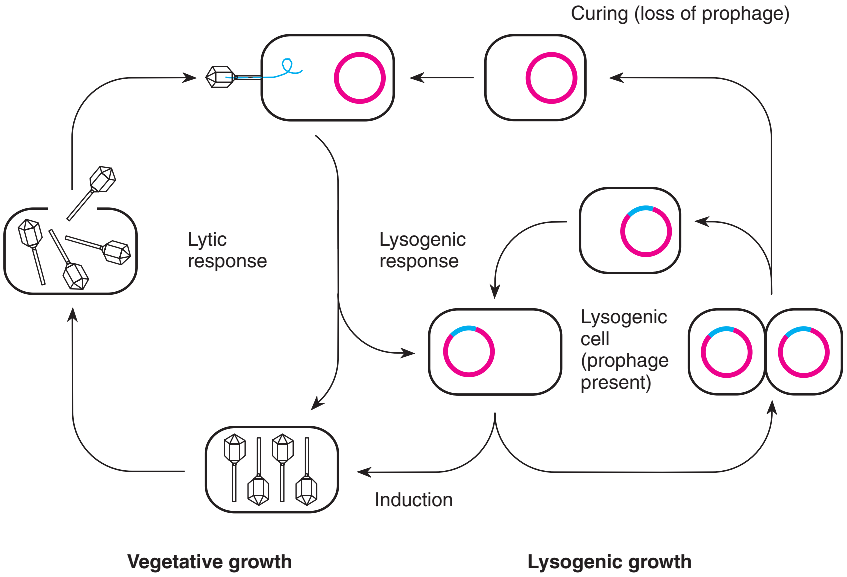


**Figure 7.24** Alternative life-cycle stages of a temperate phage (lysogenic and vegetative growth).

## *Lysogeny*

Certain phages are capable of two different life-cycle stages. Some of the time, they replicate in the host cytoplasm and destroy the host cell. At other times, these phages are capable of surviving in the host cell. The host is then referred to as **lysogenic** and the phage as **temperate.** (The term *lysogeny* means "giving birth to lysis." A lysogenic bacterium can be induced to initiate the virulent phase of the phage life cycle.)

The majority of research on lysogeny has been done on phage λ (see fig. 7.21). The λ prophage integrates into the host chromosome; other prophages, like P1, exist as independent plasmids. Phage λ, unlike the F factor, attaches at a specific point, termed *att*λ. This locus can be mapped on the *E. coli* chromosome; it lies between the galactose (*gal*) and biotin (*bio*) loci. When the phage is integrated, it protects the host from further infection (superinfection) by other λ phages. The integrated phage is now termed a **prophage.** Presumably it becomes integrated by a single crossover between itself and the host after apposition at the *att*λ site. (This process resembles the F-factor integration shown in fig. 7.18.)

A prophage can enter the lytic cycle of growth by a process of **induction,** which involves the excision of the prophage followed by the virulent or lytic stage of the viral life cycle. We consider the interesting and complex control mechanisms of life cycle in detail in chapter 14. Induction can take place through a variety of mechanisms, including UV irradiation and passage of the integrated prophage during conjugation **(zygotic induction).** The complete life cycle of a temperate phage is shown in figure 7.24.

# TRANSDUCTION

Before lysis, when phage DNA is being packaged into phage heads, an occasional error occurs that causes bacterial DNA to be incorporated into the phage head instead. When this happens, bacterial genes can be transferred to another bacterium via the phage coat. This process, called transduction, has been of great use in mapping the bacterial chromosome. Transduction occurs in two patterns: specialized and generalized.

## *Specialized Transduction*

The process of **specialized** or **restricted transduction** was first discovered in phage λ by Lederberg and his students. Specialized transduction is analogous to sexduction—it depends upon a mistake made during a looping-out process. In sexduction, the error is in the F factor. In specialized transduction, the error is in the λ prophage. Figure 7.25 shows the λ prophage looping out incorrectly to create a defective phage carrying the adjacent *gal* locus. Since only loci adjacent to the phage attachment site can be transduced in this process, specialized transduction has not proven very useful for mapping the host chromosome.

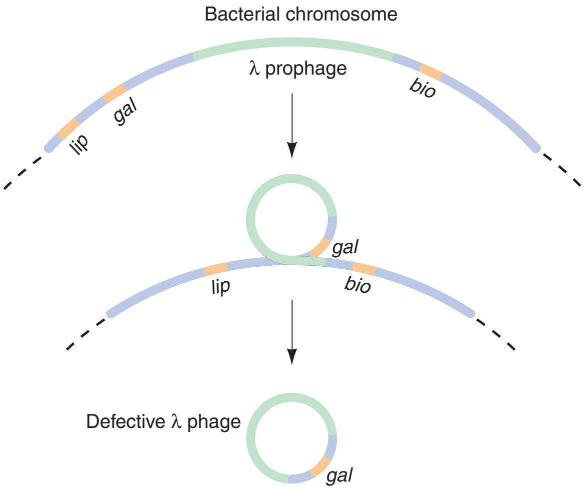


**Figure 7.25** Imprecise excision, or looping out, of the λ prophage, resulting in a defective phage carrying the *gal* locus.

## *Generalized Transduction*

**Generalized transduction,** which Zinder and Lederberg discovered, was the first mode of transduction discovered. The bacterium was *Salmonella typhimurium* and the phage was P22. Virtually any locus can be transduced by generalized transduction. The mechanism, therefore, does not depend on a faulty excision, but rather on the random inclusion of a piece of the host chromosome within the phage protein coat. A defective phage, one that carries bacterial DNA rather than phage DNA, is called a **transducing particle.** Transduction is complete when the genetic material from the transducing particle is injected into a new host and enters the new host's chromosome by recombination.

For P22, the rate of transduction is about once for every $10^5$ infecting phages. Since a transducing phage can carry only 2 to 2.5% of the host chromosome, only genes very close to each other can be transduced together **(cotransduced).** Cotransduction can thus help to fill in the details of gene order over short distances after interrupted mating or transformation is used to ascertain the general pattern. Transduction is similar to transformation in that cotransduction, like co-occurrence in transformation, is a relative indicator of map distance.

## *Mapping with Transduction*

Transduction can be used to establish gene order and map distance. Gene order can be established by two-factor transduction. For example, if gene *A* is cotransduced with gene *B* and *B* with gene *C*, but *A* is never cotransduced with *C*, we have established the order *A B C*

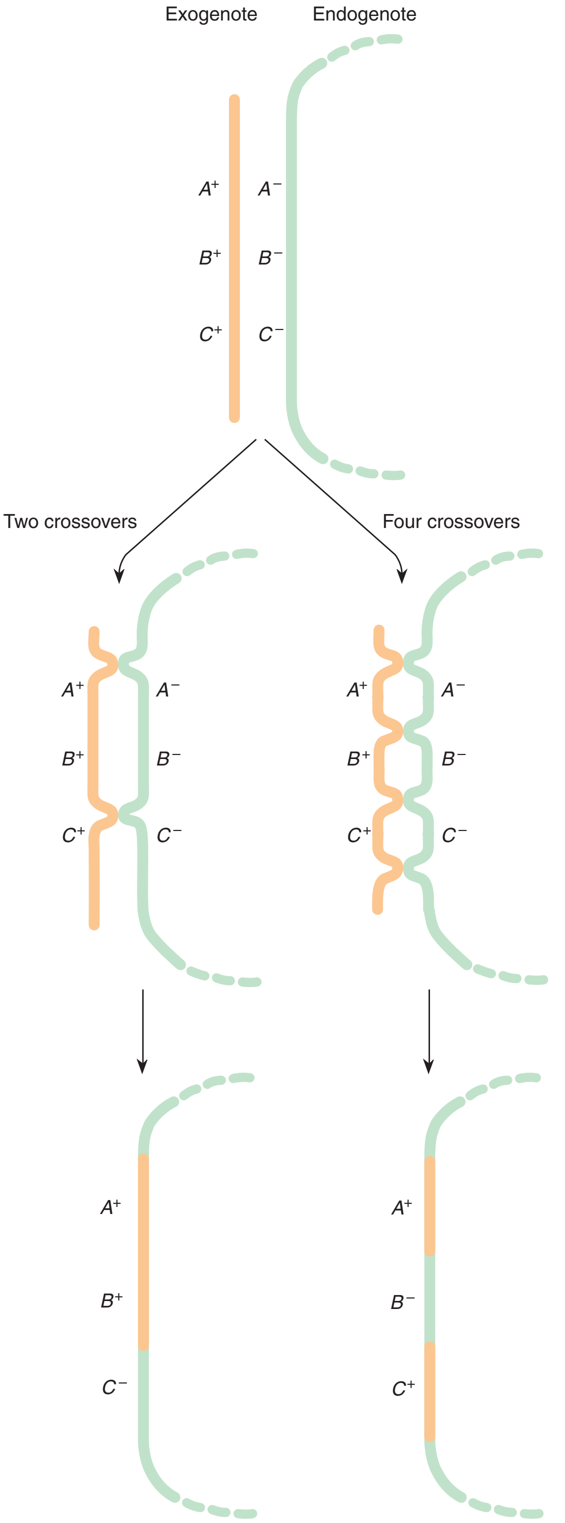


**Figure 7.26** The rarest transductant requires four crossovers.

**Table 7.5** **Gene Order Established by Two-Factor Cotransduction***

| Transductants | Number |
|---|---|
| $A^+B^+$ | 30 |
| $A^+C^+$ | 0 |
| $B^+C^+$ | 25 |
| $A^+\ B^+\ C^+$ | 0 |

* An $A^+\ B^+\ C^+$ strain of bacteria is infected with phage. The lysate is used to infect an $A^-\ B^-\ C^-$ strain. The transductants are scored for the wild-type alleles they contain. These data include only those bacteria transduced for two or more of the loci. Since *AB* cotransductants and *BC* cotransductants occur, but no *AC* types, we can infer the *A B C* order.

(table 7.5). This would also apply to quantitative differences in cotransduction. For example, if *E* is often cotransduced with *F* and *F* often with *G*, but *E* is very rarely cotransduced with *G*, then we have established the order *E F G.*

However, even more valuable is three-factor transduction, in which we can simultaneously establish gene order and relative distance. Three-factor transduction is especially valuable when the three loci are so close that it is very difficult to make ordering decisions on the basis of two-factor transduction or interrupted mating. For example, if genes *A*, *B*, and *C* are usually cotransduced, we can find the order and relative distances by taking advantage of the rarity of multiple crossovers. Let us use the prototroph ($A^+\ B^+\ C^+$) to make transducing phages that then infect the $A^-\ B^-\ C^-$ strain of bacteria.

To detect cells that have been transduced for one, two, or all three of the loci, we need to eliminate the nontransduced cells. In other words, after transduction, there will be $A^-\ B^-\ C^-$ cells in which no transduction event has taken place. There will also be seven classes of bacteria that have been transduced for one, two, or all three loci ($A^+\ B^+\ C^+$, $A^+\ B^+\ C^-$, $A^+\ B^-\ C^+$, $A^-\ B^+\ C^+$, $A^+\ B^-\ C^-$, $A^-\ B^+\ C^-$, and $A^-\ B^-\ C^+$). The simplest way to select for transduced bacteria is to select bacteria in which the wild-type has replaced at least one of the loci. For example, if, after transduction, we grow the bacteria in minimal media with the requirements of $B^-$ and $C^-$ added, all the bacteria that are $A^+$ will grow. (Without the requirement of $A^-$ bacteria, no $A^-$ bacteria will grow.) Hence, although we lose the $A^-\ B^+\ C^+$, $A^-\ B^+\ C^-$, and $A^-\ B^-\ C^+$ categories, we also lose the $A^-\ B^-\ C^-$, untransduced bacteria. In this example, the *A* locus is the selected locus; we must keep in mind that we have an incomplete, although informative, data set. Replica-plating allows us to determine genotypes at the *B* and *C* loci for the $A^+$ bacteria.

In this example, colonies that grow on complete medium without the requirement of the *A* mutant are replica-plated onto complete medium without the requirement of the *B* mutant and then onto complete medium without the requirement of the *C* mutant. In this way, each transductant can be scored for the other two loci (table 7.6). Now let us take all these selected transductants in which the $A^+$ allele was incorporated. These can be of four categories: $A^+\ B^+\ C^+$, $A^+\ B^+\ C^-$, $A^+\ B^-\ C^+$, and $A^+\ B^-\ C^-$. We can now compare the relative numbers of each of these four categories. The rarest category will be caused by the event that brings in the outer two markers, but not the center one, because this event requires four crossovers (fig. 7.26). Thus, by looking at the number of transductants in the various categories, we can determine that the gene order is *A B C* (table 7.7), since the $A^+\ B^-\ C^+$ category is the rarest.

**Table 7.6** **Method of Scoring Three-Factor Transductants**

| | Minimal Medium | | | |
|---|---|---|---|---|
| Colony Number | Without *A* Requirement | Without *B* Requirement | Without *C* Requirement | Genotype |
| 1 | + | + | − | $A^+\ B^+\ C^-$ |
| 2 | + | − | − | $A^+\ B^-\ C^-$ |
| 3 | + | − | − | $A^+\ B^-\ C^-$ |
| 4 | + | + | + | $A^+\ B^+\ C^+$ |
| 5 | + | − | + | $A^+\ B^-\ C^+$ |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |

*Note:* The plus indicates growth, the minus lack of growth. An $A^-\ B^-\ C^-$ strain was transduced by phage from an $A^+\ B^+\ C^+$ strain.

**Table 7.7** **Numbers of Transductants and Relative Cotransduction Frequencies in the Experiment Used to Determine the *A B C* Gene Order (Table 7.6)**

| Class | Number |
|---|---|
| $A^+ B^+ C^+$ | 50 |
| $A^+ B^+ C^-$ | 75 |
| $A^+ B^- C^+$ | 1 |
| $A^+ B^- C^-$ | 300 |
| | 426 |

*Relative Cotransductance*

*A-B:* (50 + 75)/426 = 0.29

*A-C:* (50 + 1)/426 = 0.12

Table 7.7 also includes calculations of the relative cotransduction frequencies. Remember that in all organisms and viruses, the higher the frequency of co-occurrence between the alleles of two loci, the closer those loci are on the chromosome. We usually measure the separation of loci by crossing over between them; the closer together, the lower those crossing-over values are and, hence, the smaller the measure of map units apart. Here, as with transformation, we are measuring the co-occurrence directly; therefore, the measure—cotransductance—is the inverse of map distance. In other words, the greater the cotransduction rate, the closer the two loci are; the more frequently two loci are transduced together, the closer they are and the higher the cotransduction value will be.

The data in table 7.7 should not be used to calculate the *B-C* cotransduction rate because the data are selected values, all of which are $A^+$; they do not encompass the total data. Missing is the $A^- B^+ C^+$ group that would contribute to the *B-C* cotransductance rate. The $A^- B^+ C^-$ and $A^- B^- C^+$ groups, also missing, would contribute only to the totals in the denominator, not the numerator, of the cotransductance index.

From these sorts of transduction experiments, it is possible to round out the details of map relations in *E. coli* after obtaining the overall picture by interrupted mating. The partial map of *E. coli* appears in figure 7.27. Definitions of loci can be found in table 7.8. Unlike the measurements in eukaryotic mapping, prokaryotic map distances are not generally thought of in map units (centimorgans). Rather, the general distance between loci is determined in *minutes* with cotransduction values used for loci that are very close to each other. (In chapter 13, we discuss mapping methods that rely on directly sequencing the DNA.)

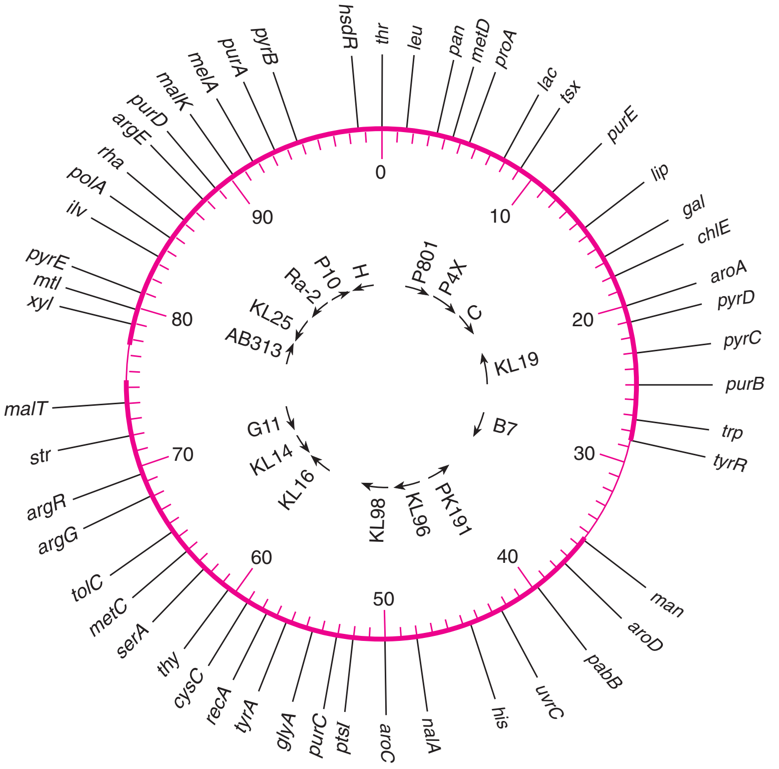

hsdR
thr
leu
pan
metD
proA
lac
tsx
purE
lip
gal
chlE
aroA
pyrD
pyrC
purB
trp
tyrR
man
aroD
pabB
uvrC
his
nalA
aroC
ptsI
purC
glyA
tyrA
recA
cysC
thy
serA
metC
tolC
argG
argR
str
malT
xyl
mtl
pyrE
ilv
polA
rha
argE
purD
malK
melA
purA
pyrB
0
10
20
30
40
50
60
70
80
90
H
P10
Ra-2
KL25
AB313
P801
P4X
C
KL19
B7
PK191
KL96
KL98
KL16
KL14
G11

**Figure 7.27** Selected loci on a circular map of *E. coli.* Definitions of loci not found in the text can be found in table 7.8. Units on the map are in minutes. *Arrows* within the circle refer to Hfr-strain transfer starting points, with directions indicated. The two thin regions on the outer circle are the only areas not covered by P1 transducing phages. (From B. J. Bachmann et al., "Recalibrated linkage map of *Escherichia coli* K-12," *Bacteriological Reviews,* 40:116–17. Copyright © 1976 American Society for Microbiology, Washington, D.C. Reprinted by permission.)

**Table 7.8** Symbols Used in the Gene Map of the *E. coli* Chromosome

| Genetic Symbols | Mutant Character | Enzyme or Reaction Affected |
|---|---|---|
| *araD* | Cannot use the sugar arabinose as a carbon source | L-Ribulose-5-phosphate-4-epimerase |
| *araA* | | L-Arabinose isomerase |
| *araB* | | L-Ribulokinase |
| *araC* | | |
| *argB* | Requires the amino acid arginine for growth | N-Acetylglutamate synthetase |
| *argC* | | N-Acetyl-γ-glutamokinase |
| *argH* | | N-Acetylglutamic-γ-semialdehyde dehydrogenase |
| *argG* | | Acetylornithine-*d*-transaminase |
| *argA* | | Acetylornithinase |
| *argD* | | Ornithine transcarbamylase |
| *argE* | | Argininosuccinic acid synthetase |
| *argF* | | Argininosuccinase |
| *argR* | Arginine operon regulator | |
| *aroA, B, C* | Requires several aromatic amino acids and vitamins for growth | Shikimic acid to 3-Enolpyruvyl-shikimate-5-phosphate |
| *aroD* | | Biosynthesis of shikimic acid |
| *azi* | Resistant to sodium azide | |
| *bio* | Requires the vitamin biotin for growth | |
| *carA* | Requires uracil and arginine | Carbamate kinase |
| *carB* | | |
| *chlA-E* | Cannot reduce chlorate | Nitrate-chlorate reductase and hydrogen lysase |
| *cysA* | Requires the amino acid cysteine for growth | 3-Phosphoadenosine-5-phosphosulfate to sulfide |
| *cysB* | | Sulfate to sulfide; four known enzymes |
| *cysC* | | |
| *dapA* | Requires the cell-wall component diaminopimelic acid | Dihydrodipicolinic acid synthetase |
| *dapB* | | N-Succinyl-diaminopimelic acid deacylase |
| *dap + hom* | Requires the amino acid precursor homoserine and the cell-wall component diaminopimelic acid for growth | Aspartic semialdehyde dehydrogenase |
| *dnaA-Z* | Mutation, DNA replication | DNA biosynthesis |
| *Dsd* | Cannot use the amino acid D-serine as a nitrogen source | D-Serine deaminase |
| *fla* | Flagella are absent | |
| *galA* | Cannot use the sugar galactose as a carbon source | Galactokinase |
| *galB* | | Galactose-1-phosphate uridyl transferase |
| *galD* | | Uridine-diphosphogalactose-4-epimerase |
| *glyA* | Requires glycine | Serine hydroxymethyl transferase |
| *gua* | Requires the purine guanine for growth | |
| *H* | The H antigen is present | |
| *his* | Requires the amino acid histidine for growth | Ten known enzymes* |
| *hsdR* | Host restriction | Endonuclease R |
| *ile* | Requires the amino acid isoleucine for growth | Threonine deaminase |
| *ilvA* | Requires the amino acids isoleucine and valine for growth | α-Hydroxy-β-keto acid rectoisomerase |
| *ilvB* | | α, β-Dihydroxyisovaleric dehydrase* |
| *ilvC* | | Transaminase B |
| *ind* (indole) | Cannot grow on tryptophan as a carbon source | Tryptophanase |
| λ (*att*λ) | Chromosomal location where prophage λ is normally inserted | |
| *lacI* | *Lac* operon regulator | |
| *lacY* | Unable to concentrate β-galactosides | Galactoside permease |
| *lacZ* | Cannot use the sugar lactose as a carbon source | β-Galactosidase |

*continued*

**Table 7.8** ***continued***

| Genetic Symbols | Mutant Character | Enzyme or Reaction Affected |
|---|---|---|
| *lacO* | Constitutive synthesis of lactose operon proteins | Defective operator |
| *leu* | Requires the amino acid leucine for growth | Three known enzymes* |
| *lip* | Requires lipoate | |
| *lon* (long form) | Filament formation and radiation sensitivity are affected | |
| *lys* | Requires the amino acid lysine for growth | Diaminopimelic acid decarboxylase |
| *lys* + *met* | Requires the amino acids lysine and methionine for growth | |
| *λrec, malT* | Resistant to phage λ and cannot use the sugar maltose | Regulator for two operons |
| *malK* | Cannot use the sugar maltose as a carbon source | Maltose permease |
| *man* | Cannot use mannose sugar | Phosphomannose isomerase |
| *melA* | Cannot use melibiose sugar | Alpha-galactosidase |
| *met A-M* | Requires the amino acid methionine for growth | Ten or more genes |
| *mtl* | Cannot use the sugar mannitol as a carbon source | Two enzymes |
| *muc* | Forms mucoid colonies | Regulation of capsular polysaccharide synthesis |
| *nalA* | Resistant to nalidixic acid | |
| *O* | The O antigen is present | |
| *pan* | Requires the vitamin pantothenic acid for growth | |
| *pabB* | Requires *p*-aminobenzoate | |
| *phe A, B* | Requires the amino acid phenylalanine for growth | |
| *pho* | Cannot use phosphate esters | Alkaline phosphatase |
| *pil* | Has filaments (pili) attached to the cell wall | |
| *plsB* | Deficient phospholipid synthesis | Glycerol 3-phosphate acyltransferase |
| *polA* | Repairs deficiencies | DNA polymerase I |
| *proA* | Requires the amino acid proline for growth | |
| *proB* | | |
| *proC* | | |
| *ptsI* | Defective phosphotransferase system | Pts-system enzyme I |
| *purA* | Requires certain purines for growth | Adenylosuccinate synthetase |
| *purB* | | Adenylosuccinase |
| *purC, E* | | 5-Aminoimidazole ribotide (AIR) to 5-aminoimidazole-4-(N-succino carboximide) ribotide |
| *purD* | | Biosynthesis of AIR |
| *pyrB* | Requires the pyrimidine uracil for growth | Aspartate transcarbamylase |
| *pyrC* | | Dihydroorotase |
| *pyrD* | | Dihydroorotic acid dehydrogenase |
| *pyrE* | | Orotidylic acid pyrophosphorylase |
| *pyrF* | | Orotidylic acid decarboxylase |
| *R gal* | Constitutive production of galactose | Repressor for enzymes involved in galactose production |
| *R1 pho, R2 pho* | Constitutive synthesis of phosphatase | Alkaline phosphatase repressor |
| *R try* | Constitutive synthesis of tryptophan | Repressor for enzymes involved in tryptophan synthesis |
| *RC* (RNA control) | Uncontrolled synthesis of RNA | |
| *recA* | Cannot repair DNA radiation damage or recombine | |
| *rhaA-D* | Cannot use the sugar rhamnose as a carbon source | Isomerase, kinase, aldolase, and regulator |
| *rpoA-D* | Problems of transcription | Subunits of RNA polymerase |
| *serA* | Requires the amino acid serine for growth | 3-Phosphoglycerate dehydrogenase |
| *serB* | | Phosphoserine phosphatase |
| *str* | Resistant to or dependent on streptomycin | |
| *suc* | Requires succinic acid | |

*continued*

**Table 7.8** ***continued***

| Genetic Symbols | Mutant Character | Enzyme or Reaction Affected |
|---|---|---|
| *supB* | Suppresses ochre mutations | t-RNA |
| *tonA* | Resistant to phages T1 and T5 (mutants called B/1, 5) | T1, T5 receptor sites absent |
| *tonB* | Resistant to phage T1 (mutants called B/1) | T1 receptor site absent |
| *T6, colK rec* | Resistant to phage T6 and colicine K | T6 and colicine receptor sites absent |
| *T4 rec* | Resistant to phage T4 (mutants called B/4) | T4 receptor site absent |
| *tsx* | T6 resistance | |
| *thi* | Requires the vitamin thiamine for growth | |
| *tolC* | Tolerance to colicine E1 | |
| *thr* | Requires the amino acid threonine for growth | |
| *thy* | Requires the pyrimidine thymine for growth | Thymidylate synthetase |
| *trpA* | Requires the amino acid tryptophan for growth | Tryptophan synthetase, A protein |
| *trpB* | | Tryptophan synthetase, B protein |
| *trpC* | | Indole-3-glycerolphosphate synthetase |
| *trpD* | | Phosphoribosyl anthranilate transferase |
| *trpE* | | Anthranilate synthetase |
| *tyrA* | Requires the amino acid tyrosine for growth | Chorismate mutase T-prephenate dehydrogenase |
| *tyrR* | | Regulates three genes |
| *uvrA-E* | Resistant to ultraviolet radiation | Ultraviolet-induced lesions in DNA are reactivated |
| *valS* | Cannot charge Valyl-tRNA | Valyl-tRNA synthetase |
| *xyl* | Cannot use the sugar xylose as a carbon source | |

*Source:* B. J. Bachmann and K. B. Low, "Linkage map of *Escherichia coli* K-12," *Microbiological Reviews,* 44:1–56. Copyright © 1990 American Society for Microbiology, Washington, D.C. Reprinted by permission.

* Denotes enzymes controlled by the homologous gene loci of *Salmonella typhimurium.*

## SUMMARY

### **STUDY OBJECTIVE 1:** To define bacteria and bacterial viruses and learn about methods of studying them 149–154

Prokaryotes (bacteria) usually have a single circular chromosome of double-stranded DNA. A bacteriophage consists of a chromosome wrapped in a protein coat. Its chromosome can be DNA or RNA. Phenotypes of bacteria include colony morphology, nutritional requirements, and drug resistance. Phage phenotypes include plaque morphology and host range. Replica-plating is a rapid screening technique for assessing the phenotype of a bacterial clone.

### **STUDY OBJECTIVE 2:** To study life cycles and sexual processes in bacteria and bacteriophages 154–166

In transformation, a competent bacterium can take up relatively large pieces of DNA from the medium. This DNA can be incorporated into the bacterial chromosome.

During the process of conjugation, the fertility factor, F, is passed from an $F^+$ to an $F^-$ cell. If the F factor integrates into the host chromosome, an Hfr cell results that can pass its entire chromosome into an $F^-$ cell. The F factor is the last region to cross into the $F^-$ cell.

In transduction, a phage protein coat containing some of the host chromosome passes to a new host bacterium. Again, recombination with this new chromosomal segment can take place.

### **STUDY OBJECTIVE 3:** To make use of the sexual processes of bacteria and their viruses to map their chromosomes 155–171

We can map the phage chromosome by measuring recombination after a bacterium has been simultaneously infected by two strains of the virus carrying different alleles. In *E. coli,* mapping is most efficiently accomplished via interrupted mating and transduction. The former provides information on general gene arrangement and the latter provides finer details.

## SOLVED PROBLEMS

**PROBLEM 1:** A wild-type strain of *B. subtilis* is transformed by DNA from a strain that cannot grow on galactose ($gal^-$) and also needs biotin for growth ($bio^-$). Transformants are isolated by exposing the transformed cells to minimal medium with penicillin, killing the wild-type cells. After the penicillin is removed, replica-plating is used to establish the genotypes of 30 transformants:

| | |
|---|---|
| Class 1 $gal^-$ $bio^-$ | 17 |
| Class 2 $gal^-$ $bio^+$ | 4 |
| Class 3 $gal^+$ $bio^-$ | 9 |

What is the relative co-occurrence of these two loci?

***Answer:*** The three classes of colonies represent the three possible transformant groups. Classes 2 and 3 are single transformants and class 1 is the double transformant. We are interested in the relative co-occurrence of the two loci. Therefore we divide the number of double transformants by the total: $r = 17/(17 + 4 + 9) = 0.57$. This is a relative value inverse to a map distance; the larger it is, the closer the loci are to each other.

**PROBLEM 2:** A $gal^-$ $bio^-$ $att\lambda^-$ strain of *E. coli* is transduced by P22 phages from a wild-type strain. Transductants are selected for by growing the cells with galactose as the sole energy source. Replica-plating and testing for lysogenic ability gives the genotypes of 106 transformants:

| | |
|---|---|
| Class 1 $gal^+$ $bio^-$ $att\lambda^-$ | 71 |
| Class 2 $gal^+$ $bio^+$ $att\lambda^-$ | 0 |
| Class 3 $gal^+$ $bio^-$ $att\lambda^+$ | 9 |
| Class 4 $gal^+$ $bio^+$ $att\lambda^+$ | 26 |

What is the gene order, and what are the relative cotransduction frequencies?

***Answer:*** We have selected all transductants that are $gal^+$. Class 2 is in the lowest frequency (0) and therefore represents the quadruple crossover between the transducing DNA and the host chromosome. From this, we see that *att*λ must be in the middle because this low-probability event is the one that would have switched only the middle locus. In other words, the two end loci would be recombinant, and the middle locus would have the host allele. We can only calculate two cotransduction frequencies because these are selected data. Note that in class 1, there is no cotransduction between *gal* and either of the other two loci; class 2 would show the cotransduction of *gal* and *bio;* class 3 represents the cotransduction of *gal* and *att*λ; and class 4 represents the cotransduction of *gal* and both other loci. Therefore, cotransduction values are

$$gal\text{–}att\lambda = (9 + 26)/106 = 35/106 = 0.33$$
$$gal\text{–}bio = (0 + 26)/106 = 26/106 = 0.25.$$

## EXERCISES AND PROBLEMS*

### BACTERIA AND BACTERIAL VIRUSES IN GENETIC RESEARCH

1. What is the nature and substance of prokaryotic chromosomes and viral chromosomes? Are viruses alive?

### TECHNIQUES OF CULTIVATION

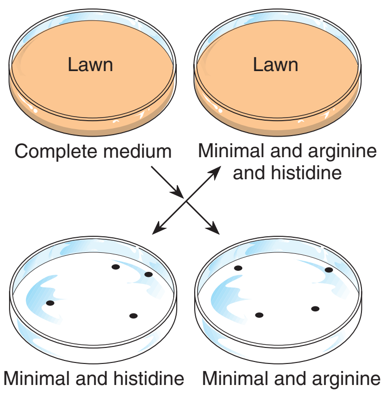
2. What are the differences between a heterotroph and an auxotroph? a minimal and a complete medium? an enriched and a selective medium?
3. What are the differences between a plaque and a colony?

### BACTERIAL PHENOTYPES

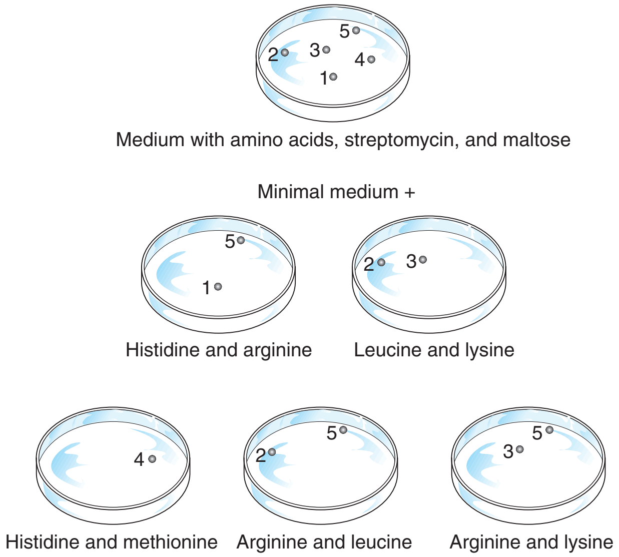
4. What genotypic notation indicates alleles that make a bacterium
   - **a.** resistant to penicillin?
   - **b.** sensitive to azide?
   - **c.** require histidine for growth?
   - **d.** unable to grow on galactose?
   - **e.** able to grow on glucose?
   - **f.** susceptible to phage T1 infection?
5. An *E. coli* cell is placed on a petri plate containing λ phages. It produces a colony overnight. By what mechanisms might it have survived?
6. An *E. coli* lawn is formed on a petri plate containing complete medium. Replica-plating is used to transfer material to plates containing minimal medium and combinations of the amino acids arginine and histidine (see the figure). Give the genotype of the original strain as well as the genotypes of the odd colonies found growing on the plates.

---

*Answers to selected exercises and problems are on page A-7.



**7.** Prototrophic Hfr *E. coli* strain G11, sensitive to streptomycin and $malT^+$ (can use maltose) is used in a conjugation experiment. The *str* locus is one of the last to be transferred, whereas the *malT* locus is one of the first. This strain is mated to an $F^-$ strain resistant to streptomycin, $malT^-$ (cannot utilize maltose), and requiring five amino acids (histidine, arginine, leucine, lysine, and methionine). Recombinants are selected for by plating on a medium with streptomycin, with maltose as the sole carbon source, and all five amino acids present. Thus, all recombinant $F^-$ cells will grow irrespective of their amino acid requirements. Five colonies are grown on the original plate with streptomycin, maltose, and all five amino acids in question (see the figure). These colonies are replica-plated onto minimal medium containing various amino acids. What are the genotypes of each of the five colonies?



**8.** A petri plate with complete medium has six colonies growing on it after one of the conjugation experiments described earlier. The colonies are numbered, and the plate is used as a master to replicate onto plates of glucose-containing selective (minimal) medium with various combinations of additives. From the following data, which show the presence (+) or absence (−) of growth, give your best assessment of the genotypes of the six colonies.

| | Colony | | | | | |
|---|---|---|---|---|---|---|
| **On Minimal Medium +** | **1** | **2** | **3** | **4** | **5** | **6** |
| Nothing | − | − | + | − | − | − |
| Xylose + arginine | + | − | + | + | − | − |
| Xylose + histidine | − | − | + | − | − | − |
| Arginine + histidine | − | + | + | + | − | − |
| Galactose + histidine | − | − | + | − | − | + |
| Threonine + isoleucine + valine | − | − | + | − | + | − |
| Threonine + valine + lactose | − | − | + | − | − | − |

## VIRAL PHENOTYPES

**9.** Give possible genotypes of an *E. coli*-phage T1 system in which the phage cannot grow on the bacterium. Give genotypes for a T1 phage that can grow on the bacterium.

## SEXUAL PROCESSES IN BACTERIA AND BACTERIOPHAGES

**10.** What is a plasmid? How does one integrate into a host's chromosome? How does it leave?

**11.** In conjugation experiments, one Hfr strain should carry a gene for some sort of sensitivity (e.g., $azi^s$ or $str^s$) so that the Hfr donors can be eliminated on selective media after conjugation has taken place. Should this locus be near to or far from the origin of transfer point of the Hfr chromosome? What are the consequences of either alternative?

**12.** How does a geneticist doing interrupted mating experiments know that the locus for the drug-sensitivity allele, used to eliminate the Hfr bacteria after conjugation, has crossed into the $F^-$ strain?

**13.** Diagram the step-by-step events required to integrate foreign DNA into a bacterial chromosome in each of the three processes outlined in the chapter (transformation, conjugation, transduction). Do the same for viral recombination. (*See also* TRANSDUCTION)

**14.** The DNA from a prototrophic strain of *E. coli* is isolated and used to transform an auxotrophic strain deficient in the synthesis of purines ($purB^-$), pyrimidines ($pyrC^-$), and the amino acid tryptophan ($trp^-$). Tryptophan was used as the marker to determine whether transformation had occurred (the selected marker). What are the gene order and the relative co-occurrence frequencies between loci, given these data:

| | |
|---|---|
| $trp^+$ $pyrC^+$ $purB^+$ | 86 |
| $trp^+$ $pyrC^+$ $purB^-$ | 4 |
| $trp^+$ $pyrC^-$ $purB^+$ | 67 |
| $trp^+$ $pyrC^-$ $purB^-$ | 14 |

**15.** Using the data in figure 7.16, draw a tentative map of the *E. coli* chromosome.

**16.** Three Hfr strains of *E. coli* (P4X, KL98, and Ra-2) are mated individually with an auxotrophic $F^-$ strain using interrupted mating techniques. Using the following data, construct a map of the *E. coli* chromosome, including distances in minutes.

| | **Approximate Time of Entry** | | |
|---|---|---|---|
| **Donor Loci** | **Hfr P4X** | **Hfr KL98** | **Hfr Ra−2** |
| $gal^+$ | 11 | 67 | 70 |
| $thr^+$ | 94 | 50 | 87 |
| $xyl^+$ | 73 | 29 | 8 |
| $lac^+$ | 2 | 58 | 79 |
| $his^+$ | 38 | 94 | 43 |
| $ilv^+$ | 77 | 33 | 4 |
| $argG^+$ | 62 | 18 | 19 |

How many different petri plates and selective media are needed?

**17.** Design an experiment using interrupted mating and create a resulting possible data set that would correctly map five of the loci on the *E. coli* chromosome (fig. 7.27).

**18.** Lederberg and his colleagues (Nester, Schafer, and Lederberg, 1963, *Genetics* 48:529) determined gene order and relative distance between genes using three markers in the bacterium *Bacillus subtilis.* DNA from a prototrophic strain ($trp^+$ $his^+$ $tyr^+$) was used to transform the auxotroph. The seven classes of transformants, with their numbers, are tabulated as follows:

| | | | | | | |
|---|---|---|---|---|---|---|
| $trp^+$ | $trp^-$ | $trp^-$ | $trp^+$ | $trp^+$ | $trp^-$ | $trp^+$ |
| $his^-$ | $his^+$ | $his^-$ | $his^+$ | $his^-$ | $his^+$ | $his^+$ |
| $tyr^-$ | $tyr^-$ | $tyr^+$ | $tyr^-$ | $tyr^+$ | $tyr^+$ | $tyr^+$ |
| 2,600 | 418 | 685 | 1,180 | 107 | 3,660 | 11,940 |

Outline the techniques used to obtain these data. Taking the loci in pairs, calculate co-occurrences. Construct the most consistent linkage map of these loci.

**19.** In a transformation experiment, an $a^+$ $b^+$ $c^+$ strain is used as the donor and an $a^-$ $b^-$ $c^-$ strain as the recipient. One hundred $a^+$ transformants are selected and then replica-plated to determine whether $b^+$ and $c^+$ are present. What can you conclude about the relative positions of the genes, based on the listed genotypes?

| | |
|---|---|
| $a^+$ $b^-$ $c^-$ | 21 |
| $a^+$ $b^-$ $c^+$ | 69 |
| $a^+$ $b^+$ $c^-$ | 3 |
| $a^+$ $b^+$ $c^+$ | 7 |

**20.** In a transformation experiment, an $a^+$ $b^+$ $c^-$ strain is used as donor and an $a^-$ $b^-$ $c^+$ strain as recipient. If you select for $a^+$ transformants, the least frequent class is $a^+$ $b^+$ $c^+$. What is the order of the genes?

**21.** A mating between $his^+$, $leu^+$, $thr^+$, $pro^+$, $str^s$ cells (Hfr) and $his^-$, $leu^-$, $thr^-$, $pro^-$, $str^r$ cells ($F^-$) is allowed to continue for twenty-five minutes. The mating is stopped, and the genotypes of the recombinants are determined. What is the first gene to enter, and what is the probable gene order, based on the following data?

| **Genotype** | **Number of Colonies** |
|---|---|
| $his^+$ | 0 |
| $leu^+$ | 12 |
| $thr^+$ | 27 |
| $pro^+$ | 6 |

**22.** **a.** In a transformation experiment, the donor is $trp^+$ $leu^+$ $arg^+$, and the recipient is $trp^-$ $leu^-$ $arg^-$. The selection process is for $trp^+$ transformants, which are then further tested. Forty percent are $trp^+$ $arg^+$; 5% are $trp^+$ $leu^+$. In what two possible orders could the genes be arranged?

**b.** You can do only one more transformation to determine gene order. You must use the same donor and recipient, but you can change the selection procedure for the initial transformants. What should you do, and what results should you expect for each order you proposed in *a*?

**23.** DNA from a bacterial strain that is $a^+$ $b^+$ $c^+$ is used to transform a strain that is $a^-$ $b^-$ $c^-$. The numbers of each transformed genotype appear. What can we say about the relative position of the genes?

| **Genotype** | | | **Number** |
|---|---|---|---|
| $a^+$ | $b^-$ | $c^-$ | 214 |
| $a^-$ | $b^+$ | $c^-$ | 231 |
| $a^-$ | $b^-$ | $c^+$ | 206 |
| $a^+$ | $b^+$ | $c^-$ | 11 |
| $a^+$ | $b^+$ | $c^+$ | 6 |
| $a^+$ | $b^-$ | $c^+$ | 93 |
| $a^-$ | $b^+$ | $c^+$ | 14 |

**24.** An Hfr strain that is $a^+\ b^+\ c^+\ d^+\ e^+$ is mated with an $F^-$ strain that is $a^-\ b^-\ c^-\ d^-\ e^-$. The mating is interrupted every five minutes, and the genotypes of the $F^-$ recombinants are determined. The results appear following. (A *plus* indicates appearance; a *minus* the lack of the locus.) Draw a map of the chromosome and indicate the position of the F factor, the direction of transfer, and the minutes between genes.

| Time | *a* | *b* | *c* | *d* | *e* |
|---|---|---|---|---|---|
| 5 | − | − | − | − | − |
| 10 | + | − | − | − | − |
| 15 | + | − | − | − | − |
| 20 | + | − | − | − | − |
| 25 | + | − | − | − | − |
| 30 | + | − | − | + | − |
| 35 | + | − | − | + | − |
| 40 | + | + | − | + | − |
| 45 | + | + | − | + | − |
| 50 | + | + | − | + | − |
| 55 | + | + | − | + | − |
| 60 | + | + | − | + | − |
| 65 | + | + | + | + | − |
| 70 | + | + | + | + | − |
| 75 | + | + | + | + | + |

**25.** A bacterial strain that is $lys^+\ his^+\ val^+$ is used as a donor, and $lys^-\ his^-\ val^-$ as the recipient. Initial transformants are isolated on minimal medium + histidine + valine.

**a.** What genotypes will grow on this medium?

**b.** These colonies are replicated to minimal medium + histidine, and 75% of the original colonies grow. What genotypes will grow on this medium?

**c.** The original colonies are also replicated to minimal medium + valine, and 6% of the colonies grow. What genotypes will grow on this medium?

**d.** Finally, the original colonies are replicated to minimal medium. No colonies grow. From this information, what genotypes will grow on minimal medium + histidine and on minimal medium + valine?

**e.** Based on this information, which gene is closer to *lys*?

**f.** The original transformation is repeated, but the original plating is on minimal medium + lysine + histidine. Fifty colonies appear. These colonies are replicated to determine their genotypes, with these results:

| | |
|---|---|
| $val^+\ his^+\ lys^+$ | 0 |
| $val^+\ his^-\ lys^+$ | 37 |
| $val^+\ his^+\ lys^-$ | 3 |

Based on all the results, what is the most likely gene order?

**LIFE CYCLES OF BACTERIOPHAGES**

**26.** Define *prophage, lysate, lysogeny,* and *temperate phage.*

**27.** Outline an experiment to demonstrate that two phages do not undergo recombination until a bacterium is infected simultaneously with both.

**28.** Doermann (1953, *Cold Spr. Harb. Symp. Quant. Biol.* 18:3) mapped three loci of phage T4: minute, rapid lysis, and turbid. He infected *E. coli* cells with both the triple mutant (*m r tu*) and the wild-type ($m^+\ r^+\ tu^+$) and obtained the following data:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $m$ | $m^+$ | $m$ | $m$ | $m^+$ | $m^+$ | $m$ | $m^+$ |
| $r$ | $r$ | $r^+$ | $r$ | $r^+$ | $r$ | $r^+$ | $r^+$ |
| $tu$ | $tu$ | $tu$ | $tu^+$ | $tu$ | $tu^+$ | $tu^+$ | $tu^+$ |
| 3,467 | 474 | 162 | 853 | 965 | 172 | 520 | 3,729 |

What is the linkage relationship among these loci? In your answer include gene order, relative distance, and coefficient of coincidence.

**29.** Wild-type phage T4 ($r^+$) produce small, turbid plaques, whereas *r*II mutants produce large, clear plaques. Four *r*II mutants (*a–d*) are crossed. (Assume, for the purposes of this problem, that *a–d* are four closely linked loci. The actual structure of the *r*II region is presented in chapter 12. Here, assume that $a \times b$ means $a^-\ b^+\ c^+\ d^+ \times a^+\ b^-\ c^+\ d^+$.) These percentages of wild-type plaques are obtained in crosses:

| | |
|---|---|
| $a \times b$ | 0.3 |
| $a \times c$ | 1.0 |
| $a \times d$ | 0.4 |
| $b \times c$ | 0.7 |
| $b \times d$ | 0.1 |
| $c \times d$ | 0.6 |

Deduce a genetic map of these four mutants.

**30.** A phage cross is performed between $a^+\ b^+\ c^+$ and $a\ b\ c$ phage. Based on these results, derive a complete map:

| | |
|---|---|
| $a^+\ b^+\ c^+$ | 1,801 |
| $a^+\ b^+\ c$ | 954 |
| $a^+\ b\ c^+$ | 371 |
| $a^+\ b\ c$ | 160 |
| $a\ b^+\ c^+$ | 178 |
| $a\ b^+\ c$ | 309 |
| $a\ b\ c^+$ | 879 |
| $a\ b\ c$ | 1,850 |
| | 6,502 |

**31.** The *r*II mutants of T4 phage will grow and produce large plaques on strain B; *r*II mutants will not grow on strain K12. Certain crosses are performed in strain B. (As with question 29, assume that the three mutants are of three separate loci in the *r*II region.) By diluting and plating on strain B, it is determined that each experiment generates about 250 × $10^7$ phage. By dilution, approximately 1/10,000 of the progeny are plated on K12 to generate these wild-type recombinants (plaques on K12):

| | |
|---|---|
| 1 × 2 | 50 |
| 1 × 3 | 25 |
| 2 × 3 | 75 |

Draw a map of these three mutants (1, 2, and 3) and indicate the distances between them.

### TRANSDUCTION

**32.** Define and illustrate *specialized* and *generalized transduction.*

**33.** In *E. coli,* the three loci *ara, leu,* and *ilvH* are within 1/2-minute map distance apart. To determine the exact order and relative distance, the prototroph (*ara*$^+$ *leu*$^+$ *ilvH*$^+$) was infected with transducing phage P1. The lysate was used to infect the auxotroph (*ara*$^-$ *leu*$^-$ *ilvH*$^-$). The *ara*$^+$ classes of transductants were selected to produce the following data:

| | | | |
|---|---|---|---|
| *ara*$^+$ | *ara*$^+$ | *ara*$^+$ | *ara*$^+$ |
| *leu*$^-$ | *leu*$^+$ | *leu*$^-$ | *leu*$^+$ |
| *ilvH*$^-$ | *ilvH*$^-$ | *ilvH*$^+$ | *ilvH*$^+$ |
| 32 | 9 | 0 | 340 |

Outline the specific techniques used to isolate the various transduced classes. What is the gene order and what are the relative cotransduction frequencies between genes? Why do some classes occur so infrequently?

**34.** Consider this portion of an *E. coli* chromosome:

*thr* *ara leu*

Three *ara* loci, *ara*-1, *ara*-2, and *ara*-3, are located in the *ara* region. A mutant of each locus (*ara*-1$^-$, *ara*-2$^-$, and *ara*-3$^-$) was isolated, and their order with respect to *thr* and *leu* was analyzed by transduction. The donor was always *thr*$^+$ *leu*$^+$ and the recipient was always *thr*$^-$ *leu*$^-$. Each *ara* mutant was used as a donor in one cross and as a recipient in another; *ara*$^+$ transductants were selected in each case. The *ara*$^+$ transductants were then scored for *leu*$^+$ and *thr*$^+$. Based on the following results, determine the order of the *ara*$^-$ mutants with respect to *thr* and *leu.*

| Cross | Recipient | Donor | Ratio: $\dfrac{thr^- \; ara^+ \; leu^+}{thr^+ \; ara^+ \; leu^-}$ |
|---|---|---|---|
| 1 | *ara*-1$^-$ | *ara*-2$^-$ | 48.5 |
| 2 | *ara*-2$^-$ | *ara*-1$^-$ | 2.4 |
| 3 | *ara*-1$^-$ | *ara*-3$^-$ | 4.0 |
| 4 | *ara*-3$^-$ | *ara*-1$^-$ | 19.1 |
| 5 | *ara*-2$^-$ | *ara*-3$^-$ | 1.5 |
| 6 | *ara*-3$^-$ | *ara*-2$^-$ | 25.5 |

**35.** An *E. coli* strain that is *leu*$^+$ *thr*$^+$ *azi*$^r$ is used as a donor in a transduction of a strain that is *leu*$^-$ *thr*$^-$ *azi*$^s$. Either *leu*$^+$ or *thr*$^+$ transductants are selected and then scored for unselected markers. The results are obtained:

| Selected Marker | Unselected Markers |
|---|---|
| *leu*$^+$ | 48% *azi*$^r$ |
| *leu*$^+$ | 2% *thr*$^+$ |
| *thr*$^+$ | 3% *leu*$^+$ |
| *thr*$^+$ | 0% *azi*$^r$ |

What is the order of the three loci?

## CRITICAL THINKING QUESTIONS

**1.** Consider the data from table 7.4. Is there another way to interpret the data other than coming from a circular bacterial chromosome?

**2.** Why might transformation have evolved, given that the bacterium is importing DNA from a dead organism?

*Suggested Readings for chapter 7 are on page B-4.*

# 8

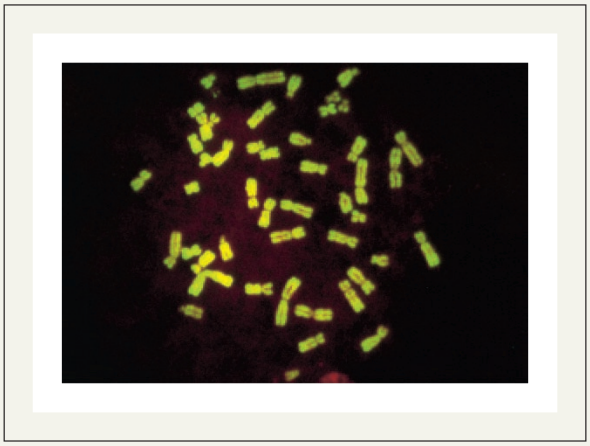
# CYTOGENETICS

## STUDY OBJECTIVES

1. To observe the nature and consequences of chromosomal breakage and reunion 178
2. To observe the nature and consequences of variation in chromosome numbers in human and nonhuman organisms 190

## STUDY OUTLINE

**Variation in Chromosomal Structure** 178
  Single Breaks 178
  Two Breaks in the Same Chromosome 179
  Two Breaks in Nonhomologous Chromosomes 182
  Centromeric Breaks 185
  Duplications 185
  Chromosomal Rearrangements in Human Beings 186
**Variation in Chromosome Number** 190
  Aneuploidy 190
  Mosaicism 190
  Aneuploidy in Human Beings 192
  Euploidy 197
*Summary* 199
*Solved Problems* 200
*Exercises and Problems* 200
*Critical Thinking Questions* 202
**Box 8.1 A Case History of the Use of Inversions to Determine Evolutionary Sequence** 182



Chromosomes of an individual with trisomy 21, Down syndrome. (© Dr. Ram Verma /Phototake, NYC.)

Our understanding of the chromosomal theory of genetics grew primarily through mapping loci, using techniques that require alternative allelic forms, or mutations, of these loci. Changes in the genetic material also occur at a much coarser level—the level of **cytogenetics,** which is a level visible under the light microscope. The word *cytogenetics* combines the words *cytology* and *genetics;* cytology is the study of cells. Cytogenetics is thus defined as the study of cells from the perspective of genetics. In practice, it is the study of changes in the gross structure and number of chromosomes in cells. In this chapter, we investigate how these alterations happen and what their consequences are to the organism.

## VARIATION IN CHROMOSOMAL STRUCTURE

In general, chromosomes can break due to ionizing radiation, physical stress, or chemical compounds. When a break occurs in the chromosome before DNA replication, during the S phase of the cell cycle (see fig. 3.6), the break itself is replicated. After the S phase, any breaks that occur affect single chromatids.

Every break in a chromatid produces two ends. These ends have been described as "sticky," meaning simply that enzymatic processes of the cell tend to reunite them. Broken ends do not attach to the undamaged terminal ends of other chromosomes.(Normal chromosomal ends are capped with structures called *telomeres*—see chapter 15.) If broken ends are not brought together, they can remain broken. But, if broken chromatid ends are brought into apposition, they may rejoin in any of several ways. First, the two broken ends of a single chromatid can reunite. Second, the broken end of one chromatid can fuse with the broken end of another chromatid, resulting in an exchange of chromosomal material and a new combination of alleles. Multiple breaks can lead to a variety of alternative recombinations. These chromosomal aberrations have major genetic, evolutionary, and medical consequences. The types of breaks and reunions discussed in this chapter can be summarized as follows:

I. Noncentromeric breaks
    A. Single breaks
        1. Restitution
        2. Deletion
        3. Dicentric bridge
    B. Two breaks (same chromosome)
        1. Deletion
        2. Inversion
    C. Two breaks (nonhomologous chromosomes)
II. Centromeric breaks
    A. Fission
    B. Fusion

### Single Breaks

If a chromosome breaks, the broken ends may rejoin. When the broken ends of a single chromatid rejoin (in a process called restitution), there is no consequence to the break. If they do not rejoin, the result is an **acentric fragment,** without a centromere, and a **centric fragment,** with a centromere. The centric fragment migrates normally during the division process because it has a centromere. The acentric fragment, however, is soon lost. It is subsequently excluded from the nuclei formed and eventually degrades. In other words, the viable, centric part of the chromosome has suffered a deletion. After mitosis, the daughter cell that receives the **deletion chromosome** may show several effects.

Pseudodominance is one possible effect. (This term was used in chapter 5 when we described alleles located on the X chromosome. With only one copy of the locus present, a recessive allele in males shows itself in the phenotype as if it were dominant—hence the term *pseudodominance.*) The normal chromosome homologous to the deletion chromosome has loci in the region, and recessive alleles show pseudodominance. A second possible effect is that, depending on the length of the deleted segment and the specific loci lost, the imbalance the deletion chromosome creates in the daughter cell may be lethal. If the deletion occurs before or during meiosis, it may be observed under the microscope. We discuss this event later in the chapter.

A single break can have yet another effect. Occasionally, the two centric fragments of a single chromosome may join, forming a two-centromere, or **dicentric, chromosome** and leaving the two acentric fragments to join or, alternatively, remain as two fragments (fig. 8.1). The acentric fragments are lost, as mentioned before. Because the centromeres are on sister chromatids, the dicentric fragment is pulled to opposite ends of a mitotic cell forming a bridge there; or, if meiosis is occurring, the dicentric fragment is pulled apart during the second meiotic division. The ultimate fate of this bridge is breakage as the spindle fibers pull the centromeres to opposite poles (or possibly exclusion from a new nucleus if the bridge is not broken).

The dicentric chromosome does not necessarily break in the middle, and subsequent processes exacerbate the imbalance created by an off-center break: duplications occur on one strand, whereas more deletions occur on the other (fig. 8.2). In addition, the "sticky" ends produced on both fragments increase the likelihood of repeating this **breakage-fusion-bridge cycle** in each generation. The great imbalances resulting from the duplications and deletions usually cause the cell line to die within several generations.

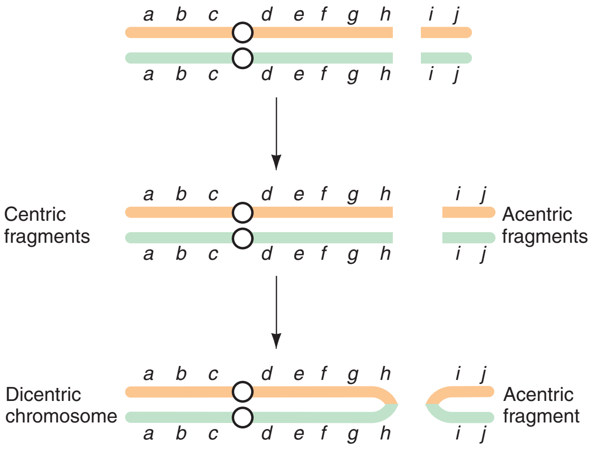


**Figure 8.1** Chromosomal break with subsequent reunion to form a dicentric chromosome and an acentric fragment.

## *Two Breaks in the Same Chromosome*

### *Deletion*

Figure 8.3 shows two of the possible results when two breaks occur in the same chromosome. One alternative is a reunion that omits an acentric fragment, which is then lost. The centric piece, missing the acentric fragment (*e-f-g* in fig. 8.3), is a deletion chromosome. An organism having this chromosome and a normal homologue will have, during meiosis, a bulge in the tetrad if the deleted section is large enough (fig. 8.4). The bulge also appears in the paired, polytene giant salivary gland chromosomes of *Drosophila.* (Note that when a bulge like that illustrated in figure 8.4 is seen in paired chromosomes, it indicates that one chromosome has a piece that is missing in the other. In our example, the bulge resulted from a deletion in one chromosome; it could also result from an insertion of a piece in the other chromosome.)

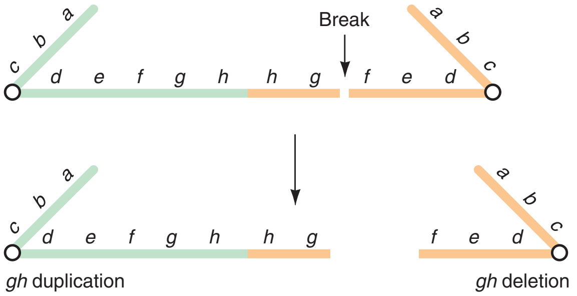


**Figure 8.2** Breakage of a dicentric bridge causes duplications and additional deficiencies.

### *Inversion*

Two breaks in the same chromosome can also lead to **inversion,** in which the middle section is reattached but in the inverted configuration (see fig. 8.3). An inversion has several interesting properties. To begin with, fruit flies homozygous for an inversion show new linkage relations when their chromosomes are mapped. One outcome of this new linkage arrangement is the possibility of a **position effect,** a change in the expression of a gene due to a changed linkage arrangement. Position effects are either stable, as in *Bar* eye of *Drosophila* (to be discussed), or variegated, as with *Drosophila* eye color. A normal female fly that is heterozygous ($X^wX^+$) has red eyes. If, however, the white locus is moved through an inversion so that it comes to lie next to heterochromatin (fig. 8.5), the fly shows a **variegation**—patches of the eye are white. This is presumably caused by a spread of the tight coiling of the heterochromatin, "turning off" the expression of the locus. In a heterozygote, if the turned-off allele is the wild-type, the cell will express the normally recessive white-eye allele. Depending on what happens in each cell, patches of red and white eye color result.

When synapsis occurs in an inversion heterozygote, either at meiosis or in the *Drosophila* salivary gland during endomitosis, a loop often forms to accommodate the point-for-point pairing process (figs. 8.6 and 8.7). An outcome of this looping tendency is **crossover suppression.** That is, an inversion heterozygote shows very little recombination of alleles within the inverted region. The reason is usually not that crossing over is actually

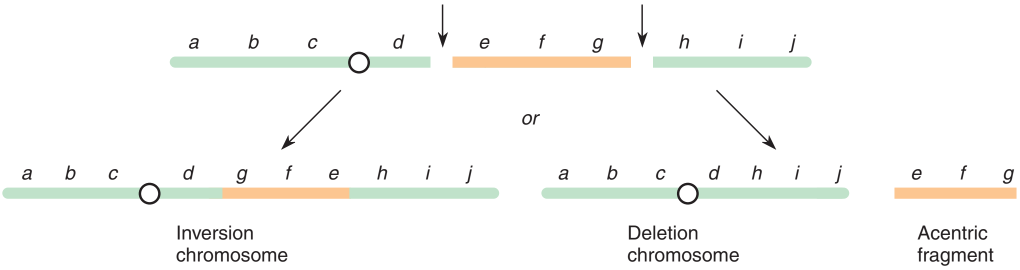


**Figure 8.3** Two possible consequences of a double break (*top arrows*) in the same chromosome.

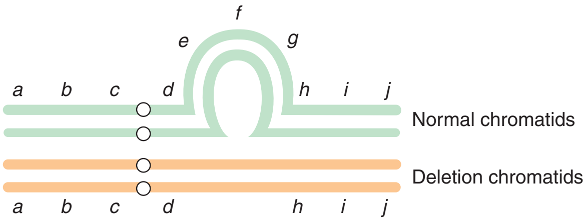


**Figure 8.4** A bulge can occur in a meiotic tetrad if a large deletion has occurred.

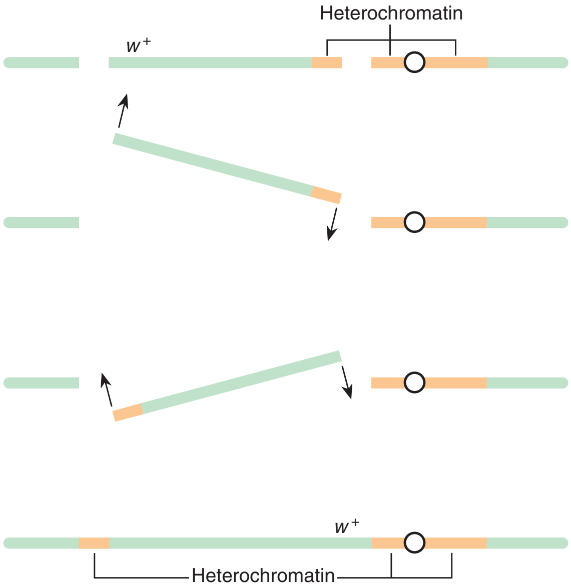


**Figure 8.5** An inversion in the X chromosome of *Drosophila* produces a variegation in eye color in a female if her other chromosome is normal and carries the white-eye allele ($X^w$).

suppressed, but rather that the products of recombination within a loop are usually lost. (Suppression can also occur in small inversions where loops don't form.) Figure 8.8 shows a crossover within a loop. The two nonsister chromatids not involved in a crossover in the loop will end up in normal gametes (carrying either the normal chromosome or the intact inverted chromosome). The products of the crossover, rather than being a simple recombination of alleles, are a dicentric and an acentric chromatid. The acentric chromatid is not incorporated into a gamete nucleus, whereas the dicentric chromatid begins a breakage-fusion-bridge cycle that creates a genetic imbalance in the gametes. The gametes thus carry chromosomes with duplications and deficiencies.

The inversion pictured in figure 8.8 is a **paracentric inversion,** one in which the centromere is outside the inversion loop. A **pericentric inversion** is one in which the inverted section contains the centromere. It, too, suppresses crossovers, but for slightly different reasons (fig. 8.9). All four chromatid products of a single crossover within the loop have centromeres and are thus incorporated into the nuclei of gametes. However, the two recombinant chromatids are unbalanced—they both have duplications and deficiencies. One has a duplication for

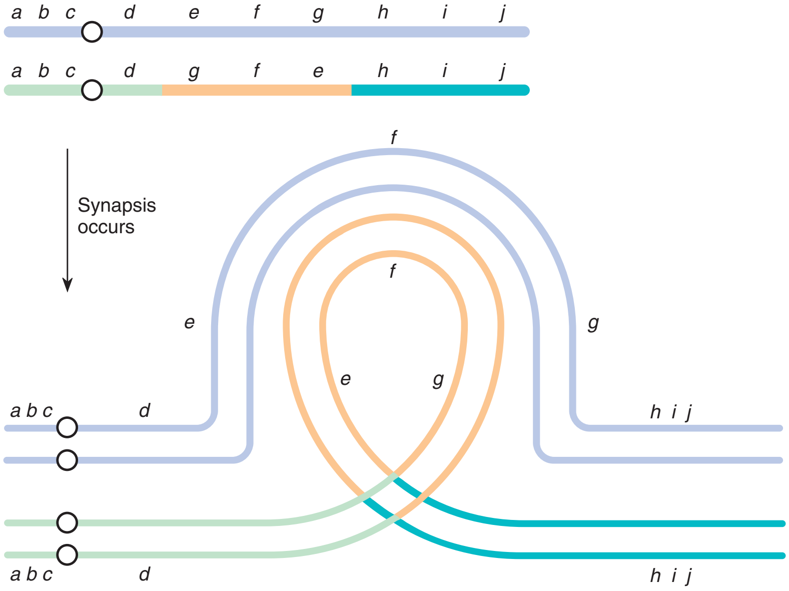


**Figure 8.6** Tetrad at meiosis showing the loop characteristic of an inversion heterozygote.

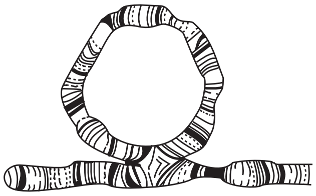


**Figure 8.7** A *Drosophila* heterozygous for an inversion will show a loop in the salivary gland chromosomes. (Compare with figure 8.6.)

*a-b-c-d* and is deficient for *h-i-j,* whereas the other is the reciprocal—deficient for *a-b-c-d* and duplicated for *h-i-j* (in fig. 8.9). These duplication-deletion gametes tend to form inviable zygotes. The result, as with the paracentric inversion, is the apparent suppression of crossing over.

### *Results of Inversion*

Crossing over within inversion loops results in **semisterility.** Almost all gametes that contain dicentric or imbalanced chromosomes form inviable zygotes. Thus, a certain proportion of the progeny of inversion heterozygotes are not viable.

Inversions have several evolutionary ramifications. Those alleles originally together in the noninversion chromosome and those found together within the inversion loop tend to stay together because of the low rate of successful recombination within the inverted region. If several loci affect the same trait, the alleles are referred to as a **supergene.** Until careful genetic analysis is done, the loci in a supergene could be mistaken for a single locus; they affect the same trait and are inherited apparently as a single unit. Examples include shell color and pattern in land snails and mimicry in butterflies (see chapter 21). Supergenes can be beneficial when they involve favorable gene combinations. However, at the same time, their inversion structure prevents the formation of new complexes. Supergenes, therefore, have evolutionary advantages and disadvantages. Chapter 21 discusses these evolutionary topics in more detail.

Sometimes the inversion process produces a record of the evolutionary history of a group of species. As species evolve, inversions can occur on preexisting inversions. This leads to very complex arrangements of loci. We can readily study these patterns in Diptera by noting the changed patterns of bands in salivary gland chromosomes. Since certain arrangements can only come about by a specific sequence of inversions, it is possible to know which species evolved from which. The same series of events can occur within the same species (box 8.1).

In summary then, inversions result in suppressed crossing over, semisterility, variegation position effects, and new linkage arrangements. All of these events have evolutionary consequences.

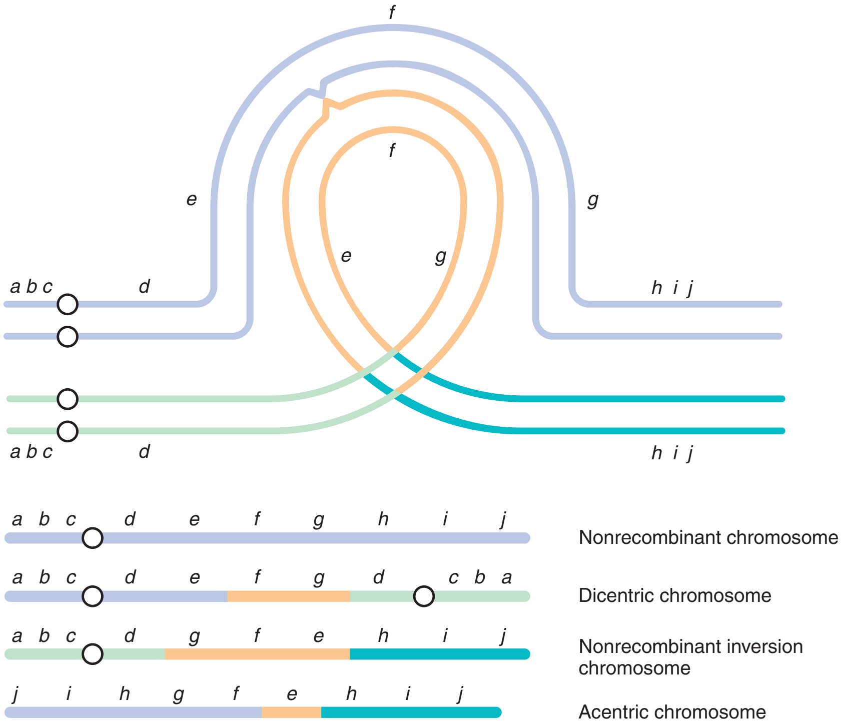


**Figure 8.8** Consequences of a crossover in the loop region of a paracentric inversion heterozygote.

## BOX 8.1

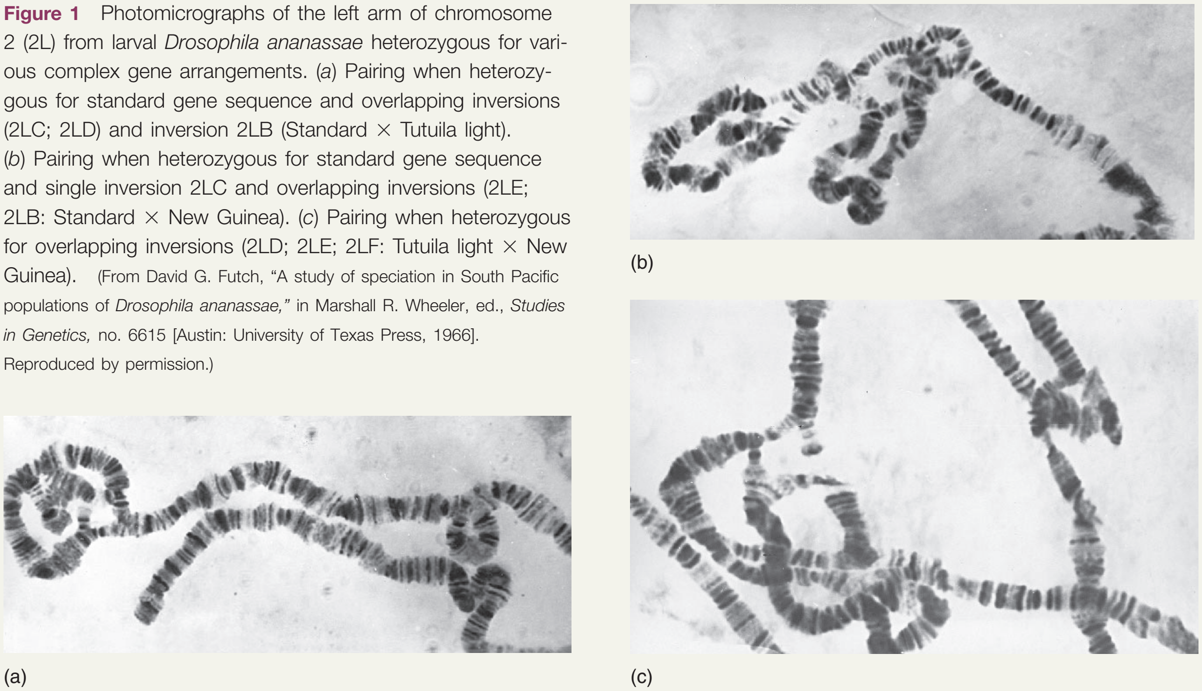
## Experimental Methods

### *A Case History of the Use of Inversions to Determine Evolutionary Sequence*

In 1966, David Futch published a study of the chromosomes of a fruit fly, *Drosophila ananassae,* an organism widely distributed throughout the tropical Pacific. The study was designed to determine something about the species status of various melanic forms of the fly. In the course of his work, Futch looked at the salivary gland chromosomes of flies from twelve different localities. He discovered twelve paracentric inversions, three pericentric inversions, and one translocation. Because of the precise banding patterns of these chromosomes, it was possible to determine the breakage points for each inversion.

Observation of several populations that have had sequential changes in their chromosomes makes it possible to determine the sequence of successive changes. Once one knows the sequence of changes in different populations of *Drosophila ananassae,* along with the geographic locations of the populations, it is possible to determine the history of the way the flies colonized these tropical islands. *D. ananassae* is particularly suited to this type of work because it is believed to be a recent invader to most of the Pacific Islands that it occupies. It is of interest to know about the spread of this species as an adjunct to studies of human migration in the Pacific Islands

**Figure 1** Photomicrographs of the left arm of chromosome 2 (2L) from larval *Drosophila ananassae* heterozygous for various complex gene arrangements. (*a*) Pairing when heterozygous for standard gene sequence and overlapping inversions (2LC; 2LD) and inversion 2LB (Standard × Tutuila light). (*b*) Pairing when heterozygous for standard gene sequence and single inversion 2LC and overlapping inversions (2LE; 2LB: Standard × New Guinea). (*c*) Pairing when heterozygous for overlapping inversions (2LD; 2LE; 2LF: Tutuila light × New Guinea). (From David G. Futch, "A study of speciation in South Pacific populations of *Drosophila ananassae,"* in Marshall R. Wheeler, ed., *Studies in Genetics,* no. 6615 [Austin: University of Texas Press, 1966]. Reproduced by permission.)



## *Two Breaks in Nonhomologous Chromosomes*

Breaks can occur simultaneously in two nonhomologous chromosomes. Reunion can then take place in various ways. The most interesting case occurs when the ends of two nonhomologous chromosomes are translocated to each other in a **reciprocal translocation** (fig. 8.10). The organism in which this has happened, a reciprocal translocation heterozygote, has all the genetic material of the normal homozygote. Two outcomes of a reciprocal translocation, like those of an inversion, are new linkage arrangements in a homozygote—an organism with translocated chromosomes only—and variegation position effects.

During synapsis, either at meiosis or endomitosis, a point-for-point pairing in the translocation heterozygote

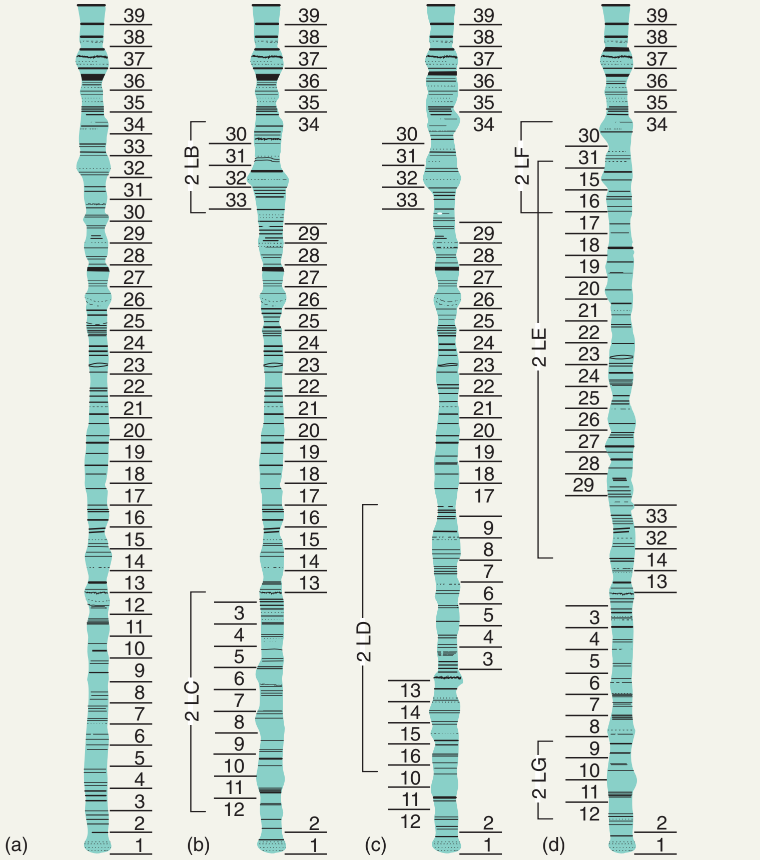


**Figure 2** Chromosomal maps of 2L. (*a*) Standard gene sequence. (*b*) Ponape: breakpoints of 2LC and 2LB are indicated and the segments are shown inverted. (*c*) Tutuila light: breakpoints of 2LD are indicated. 2LC and 2LB are inverted. 2LD, which overlaps 2LC, is also shown inverted. (*d*) New Guinea: breakpoints of 2LE and 2LG are indicated. 2LC, 2LB, and 2LE are shown inverted. Note: only the breakpoints of 2LF and 2LG are shown; neither of these is inverted in the map. (From David G. Futch, "A study of speciation in South Pacific populations of *Drosophila ananassae,*" in Marshall R. Wheeler, ed., *Studies in Genetics,* no. 6615 [Austin: University of Texas Press, 1966]. Reproduced by permission.)

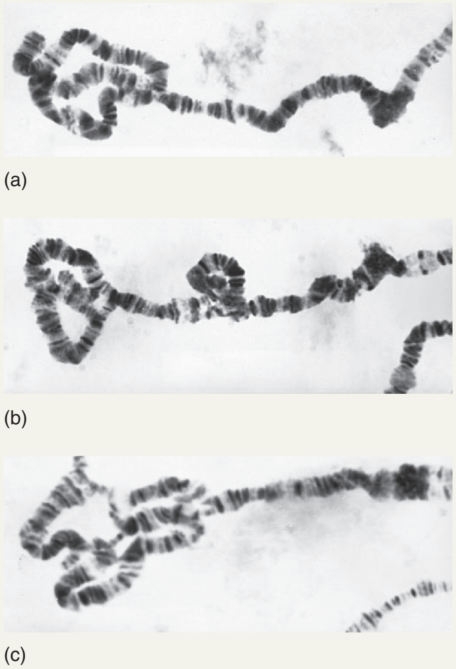

(a)
(b)
(c)

**Figure 3** Photomicrographs of the right arm of chromosome 2 (2R) from larvae heterozygous for various complex gene arrangements. (*a*) Pairing when heterozygous for standard gene sequence and overlapping inversions (2RA; 2RB: Standard × Tutuila light). (*b*) Pairing when heterozygous for standard gene sequence and overlapping inversions (2RA; 2RC) and inversion 2RD (Standard × New Guinea). (*c*) Pairing when heterozygous for overlapping inversions 2RB, 2RC, and 2RD. Inversion 2RA is homozygous (Tutuila light × New Guinea). (From David G. Futch, "A study of speciation in South Pacific populations of *Drosophila ananassae,*" in Marshall R. Wheeler, ed., *Studies in Genetics,* no. 6615 [Austin: University of Texas Press, 1966]. Reproduced by permission.)

*continued*

can be accomplished by the formation of a cross-shaped figure (fig. 8.10). Such a figure is diagnostic of a reciprocal translocation. A single crossover in a reciprocal translocation heterozygote will not produce chromatids that are further imbalanced, as it does in an inversion heterozygote. However, reciprocal translocation heterozygotes do produce nonviable progeny. Problems can arise when centromeres separate at the first meiotic division.

### *Segregation After Translocation*

Since two homologous pairs of chromosomes are involved, we have to keep track of the independent

## BOX 8.1 CONTINUED

because *D. ananassae* is commensal with people.

Some of Futch's results are shown in figures 1–4, which diagram the left and right arms of the fly's second chromosome, as well as the synaptic patterns. We can see vividly the sequence of change in which one inversion occurs after a previous inversion has already taken place. In figures 2 and 4, the standard (*a*) gave rise to (*b*), which then gave rise independently to (*c*) and (*d*). The standard is from Majuro in the Marshall Islands and is believed to be in the ancestral group of the species. Ponape is the home of (*b*), (*c*) is from Tutuila (eastern Samoa), and (*d*) is from New Guinea. Thus, the sequence is Majuro to Ponape, and from there the same stock was transferred to Tutuila and New Guinea. This type of analysis has been useful in the *Drosophila* group throughout its range but especially in the Pacific Island populations and in the southwestern United States.

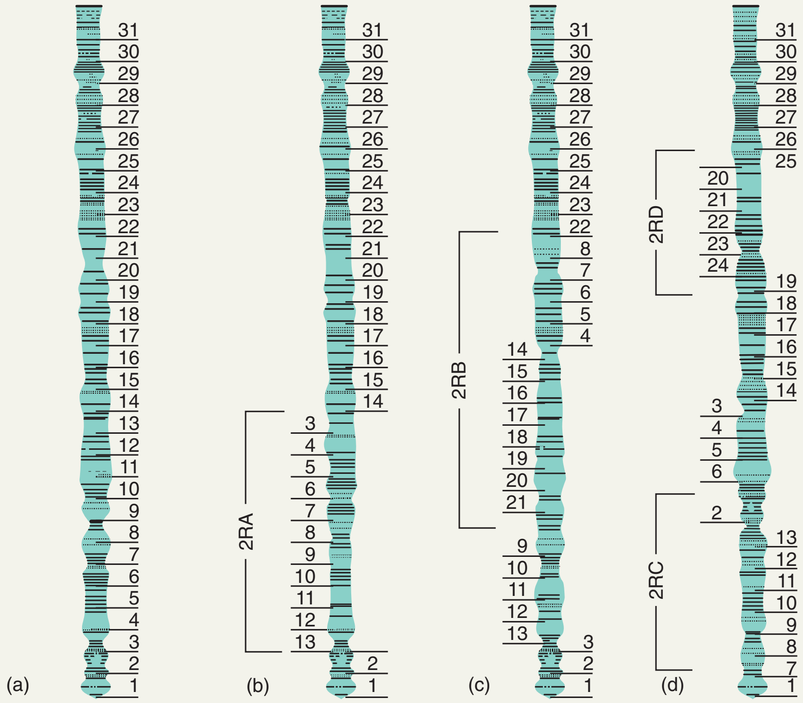

31
30
29
28
27
26
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
(a)
31
30
29
28
27
26
25
24
23
22
21
20
19
18
17
16
15
14
3
4
5
6
7
8
9
10
11
12
13
2
1
2RA
(b)
31
30
29
28
27
26
25
24
23
22
8
7
6
5
4
14
15
16
17
18
19
20
21
9
10
11
12
13
3
2
1
2RB
(c)
31
30
29
28
27
26
25
20
21
22
23
24
19
18
17
16
15
14
3
4
5
6
2
13
12
11
10
9
8
7
1
2RD
2RC
(d)

**Figure 4** Chromosomal maps of 2R. (*a*) Standard gene sequence. (*b*) Ponape: breakpoints of 2RA are indicated and the segment is shown inverted. (*c*) Tutuila light: breakpoints of 2RB are indicated. 2RA is inverted and 2RB, which overlaps it, is also shown inverted. (*d*) New Guinea: breakpoints of 2RC and 2RD are indicated. 2RA is inverted; 2RC, which overlaps 2RA, and 2RD are shown inverted. (From David G. Futch, "A study of speciation in South Pacific populations of *Drosophila ananassae,"* in Marshall R. Wheeler, ed., *Studies in Genetics,* no. 6615 [Austin: University of Texas Press, 1966]. Reproduced by permission.)

segregation of the centromeres of the two tetrads. There are two common possibilities and one that occurs less often (fig. 8.11). The first, called **alternate segregation,** occurs when the first centromere assorts with the fourth centromere, leaving the second and third centromeres to go to the opposite pole. The result will be balanced gametes, one with normal chromosomes and the other with a reciprocal translocation. Also likely is the **adjacent-1** type of segregation, in which the first and third centromeres segregate together in the opposite direction from the second and fourth centromeres. Here, both types of gametes are unbalanced, carrying duplica-

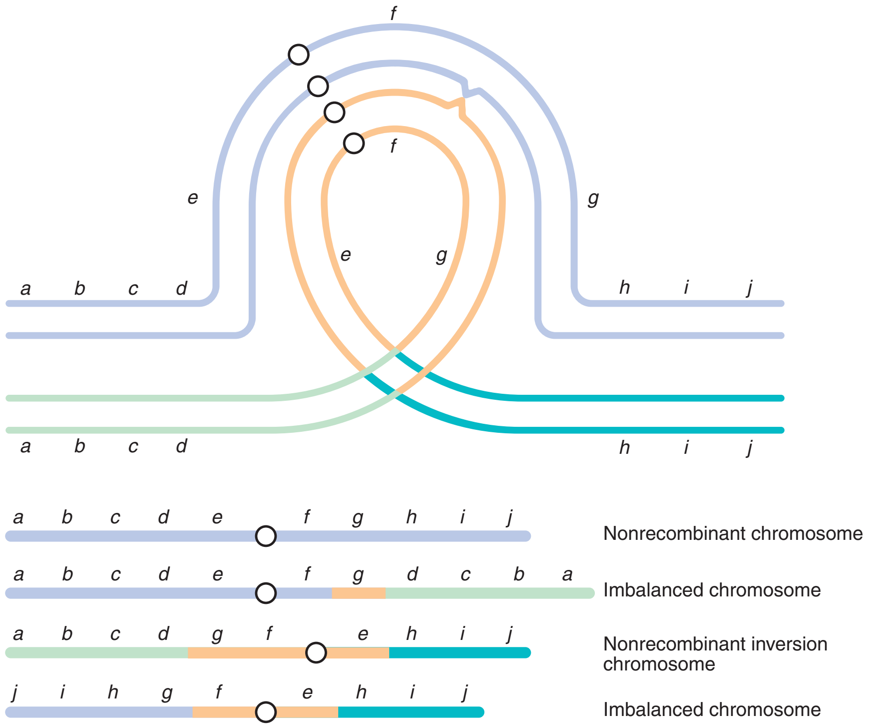


**Figure 8.9** Consequences of a crossover in the loop region of a pericentric inversion heterozygote.

tions and deficiencies that are usually lethal. Since adjacent-1 segregation occurs at a relatively high frequency, a significant amount of sterility results from the translocation (as much as 50%).

An **adjacent-2** type of segregation (fig. 8.11), in which homologous centromeres go to the same pole (first with second, third with fourth), is a third possibility. This can result when the cross-shaped double tetrad opens into a circle in late prophase I. In the German cockroach, adjacent-2 patterns have been observed in 10 to 25% of meioses, depending upon which chromosomes are involved.

In summary, then, reciprocal translocations result in new linkage arrangements, variegated position effects, a cross-shaped figure during synapsis, and semisterility.

## Centromeric Breaks

Another interesting variant of the simple reciprocal translocation occurs when two acrocentric chromosomes join at or very near their centromeres. The process, called a **Robertsonian fusion** after cytologist W. Robertson, produces a decrease in the number of chromosomes, although virtually the same amount of genetic material is maintained. Often, closely related species undergo Robertsonian fusions and end up with markedly different chromosome numbers without any significant difference in the quantity of their genetic material. Therefore, cytologists frequently count the number of chromosomal arms rather than the number of chromosomes to get a more accurate picture of species affinities. The number of arms is referred to as the **fundamental number,** or **NF** (French: *nombre fondamentale*). In a similar fashion, **centromeric fission** increases the chromosome number without changing the fundamental number.

## Duplications

Duplications of chromosomal segments can occur, as we have just seen, by the breakage-fusion-bridge cycle or by crossovers within the loop of an inversion. There is another way that duplications arise in small adjacent regions of a chromosome. We illustrate this with a particularly interesting example, the *Bar* eye phenotype in *Drosophila* (fig. 8.12). The wild-type fruit fly has about 800 facets in each eye. The *Bar* (*B*) homozygote has about 70 (a range of 20–120 facets). Another allele, *Doublebar* (*BB:* sometimes referred to as *Ultrabar,* $B^U$), brings the facet number of the eye down to about 45 when heterozygous and to about 25 when homozygous. Around 1920, researchers

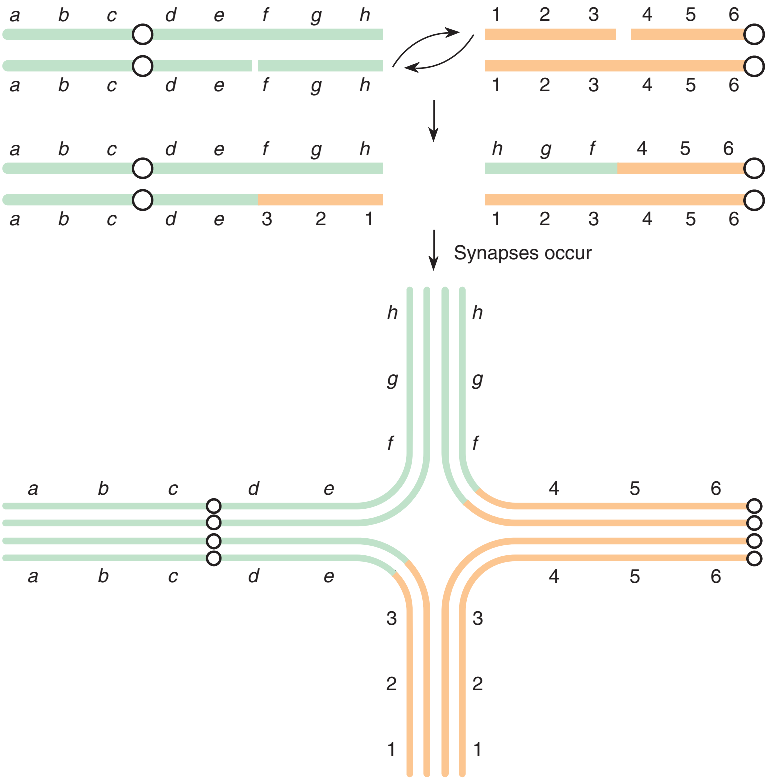


**Figure 8.10** A reciprocal translocation heterozygote forms after breaks occur in nonhomologous chromosomes. Synapsis at meiosis forms a cross-shaped figure.

showed that about one progeny in 1,600 from homozygous *Bar* females is *Doublebar.* This is much more frequent than we expect from mutation.

Alfred Sturtevant found that in every *Doublebar* fly, a crossover had occurred between loci on either side of the *Bar* locus. He suggested that the change to *Doublebar* was due to **unequal crossing over** rather than to a simple mutation of one allele to another (fig. 8.13). If the homologous chromosomes do not line up exactly during synapsis, a crossover produces an unequal distribution of chromosomal material. Later, an analysis of the banding pattern of the salivary glands confirmed Sturtevant's hypothesis. It was found that *Bar* is a duplication of several bands in the 16A region of the X chromosome (fig. 8.14). *Doublebar* is a triplication of the segment.

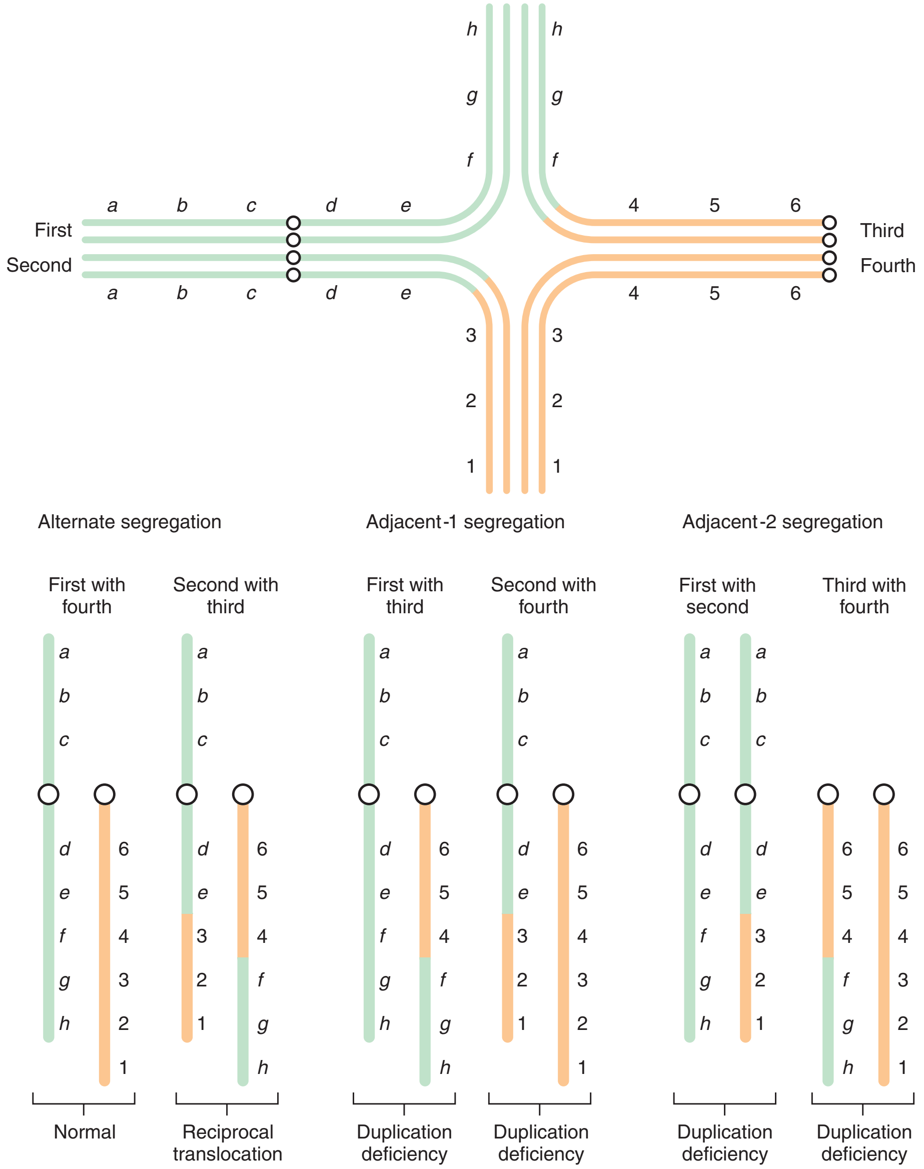
A position effect also occurs in the *Bar* system. A *Bar* homozygote (*B*/*B*) and a *Doublebar*/wild-type heterozygote ($BB/B^{+}$) both have four copies of the 16A region. It would therefore be reasonable to expect that both genotypes would produce the same phenotype. However, the *Bar* homozygote has about seventy facets in each eye, whereas the heterozygote only has about forty-five. Thus, not only the amount of genetic material, but also its configuration, determines the extent of the phenotype. *Bar* eye was the first position effect discovered.

## *Chromosomal Rearrangements in Human Beings*

Several human syndromes and abnormalities are the result of chromosomal rearrangements, including deletions and translocations. The most common are described here. Keep three points in mind as you read. First, all of these disorders are rare. Second, the deletion syndromes are often caused by a balanced translocation in one of the parents. And third, about one in five hundred live births contains a balanced rearrangement of some kind, either a reciprocal translocation or inversion.

### *Fragile-X Syndrome*

The most common cause of inherited mental retardation is the **fragile-X syndrome.** It occurs in about one in every 1,250 males and about one in every 2,000 females. Symptoms include mental retardation, altered speech patterns, and other physical attributes. The condition is



h
g
f
a b c d e
4 5 6
First
Second
Third
Fourth
3
2
1
Alternate segregation
Adjacent-1 segregation
Adjacent-2 segregation
First with fourth
Second with third
First with third
Second with fourth
First with second
Third with fourth
Normal
Reciprocal translocation
Duplication deficiency
Duplication deficiency
Duplication deficiency
Duplication deficiency

**Figure 8.11** Three possible results of chromatid separation during meiosis in a reciprocal translocation heterozygote.

called the fragile-X syndrome because it is related to a region at the X chromosome tip that breaks more frequently than other chromosomal regions. However, the break is not required for the syndrome to occur, and the fragile-X chromosome is usually identified by the lack of chromatin condensation at the site; in fact, under the microscope, it appears that the tip of the chromosome is being held in place by a thread (fig. 8.15). The gene responsible for the syndrome is called *FMR-1,* for fragile-X mental retardation-1.

Fragile-X syndrome has a highly unusual pattern of inheritance: the chance of inheriting the disease increases through generations. This is so unusual a pattern that it was termed the Sherman Paradox. Approximately 20% of males with the fragile-X chromosome do not have symptoms but have grandchildren who do have the symptoms. The daughters of the symptomatic males also don't have symptoms, but obviously, they have another X chromosome to mask the symptoms. As generations proceed, the percentage of affected sons of

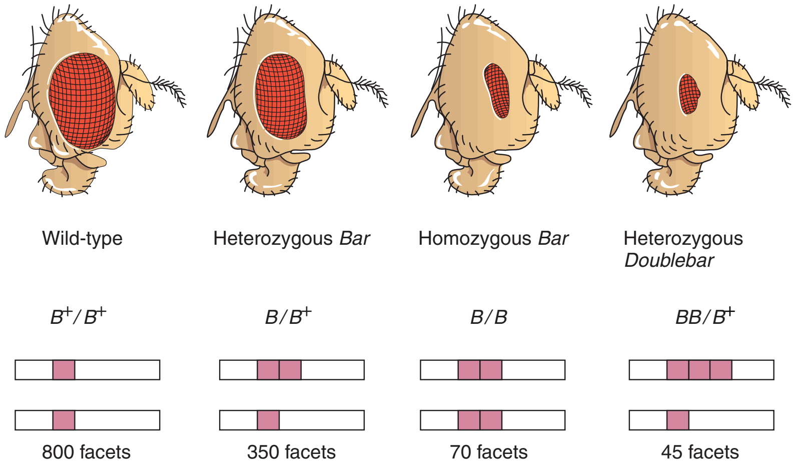

Wild-type
Heterozygous *Bar*
Homozygous *Bar*
Heterozygous *Doublebar*
$B^+/B^+$
$B/B^+$
$B/B$
$BB/B^+$
800 facets
350 facets
70 facets
45 facets

**Figure 8.12** *Bar* eye in *Drosophila* females.

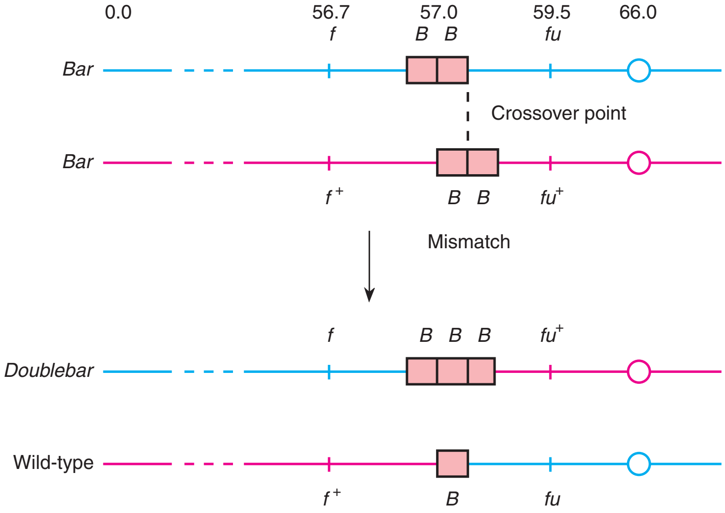

0.0
56.7
*f*
57.0
*B B*
59.5
*fu*
66.0
*Bar*
Crossover point
*Bar*
$f^+$
*B B*
$fu^+$
Mismatch
*f*
*B B B*
$fu^+$
*Doublebar*
Wild-type
$f^+$
*B*
*fu*

**Figure 8.13** Unequal crossing over in a female *Bar*-eyed *Drosophila* homozygote as a result of improper pairing. A *Doublebar* chromosome (and concomitant wild-type chromosome) is produced by a crossover between forked (*f*) and fused (*fu*), two flanking loci.

carrier mothers increases. Molecular techniques, discussed in chapter 13, revealed the odd nature of this syndrome.

Basically, the *FMR-1* gene normally has between 6 and 50 copies of a three-nucleotide repeat, CCG. Chromosomes that have the fragile-site appearance have between 230 and 2,000 copies of the repeat. The number of repeats is very unstable; when carrier women transmit the chromosome, the number of repeats usually goes up. Repeat numbers above 230 inactivate the gene and thus cause the syndrome in men, who have only one copy of the X chromosome. The function of the gene is not currently known. This unusual form of inheritance, with unstable repeats in a gene, seems to be the mechanism in several other diseases as well, including muscular dystrophy and Huntington disease. We will discuss other unusual modes of inheritance in chapter 17.

### Cri du Chat Syndrome, 46,XX or XY,5p–

The syndrome known as *cri du chat* (French: cry of the cat) is so called because of the catlike cry that about half the affected infants make. Microcephaly (an abnormally small head), congenital heart disease, and severe mental retardation are also common symptoms. This disorder arises from a deletion in chromosome 5 (fig. 8.16); most

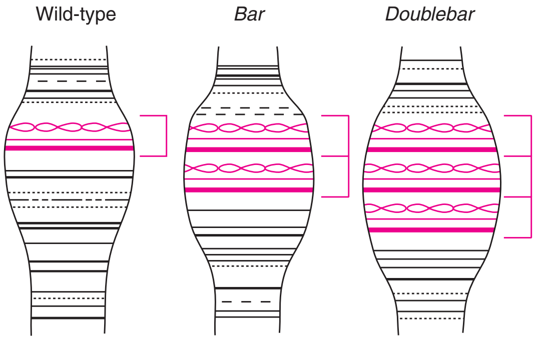

Wild-type
*Bar*
*Doublebar*

**Figure 8.14** *Bar* region of the X chromosome of *Drosophila.*

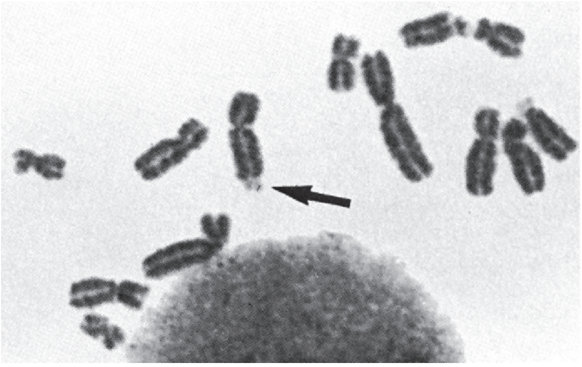


**Figure 8.15** Human metaphase chromosomes with the fragile-X site indicated by an arrow. (From Ian Craig, "Methylation and the Fragile X," *Nature* [1991] 349:742. Copyright © 1991 Macmillan Magazines, Ltd.)

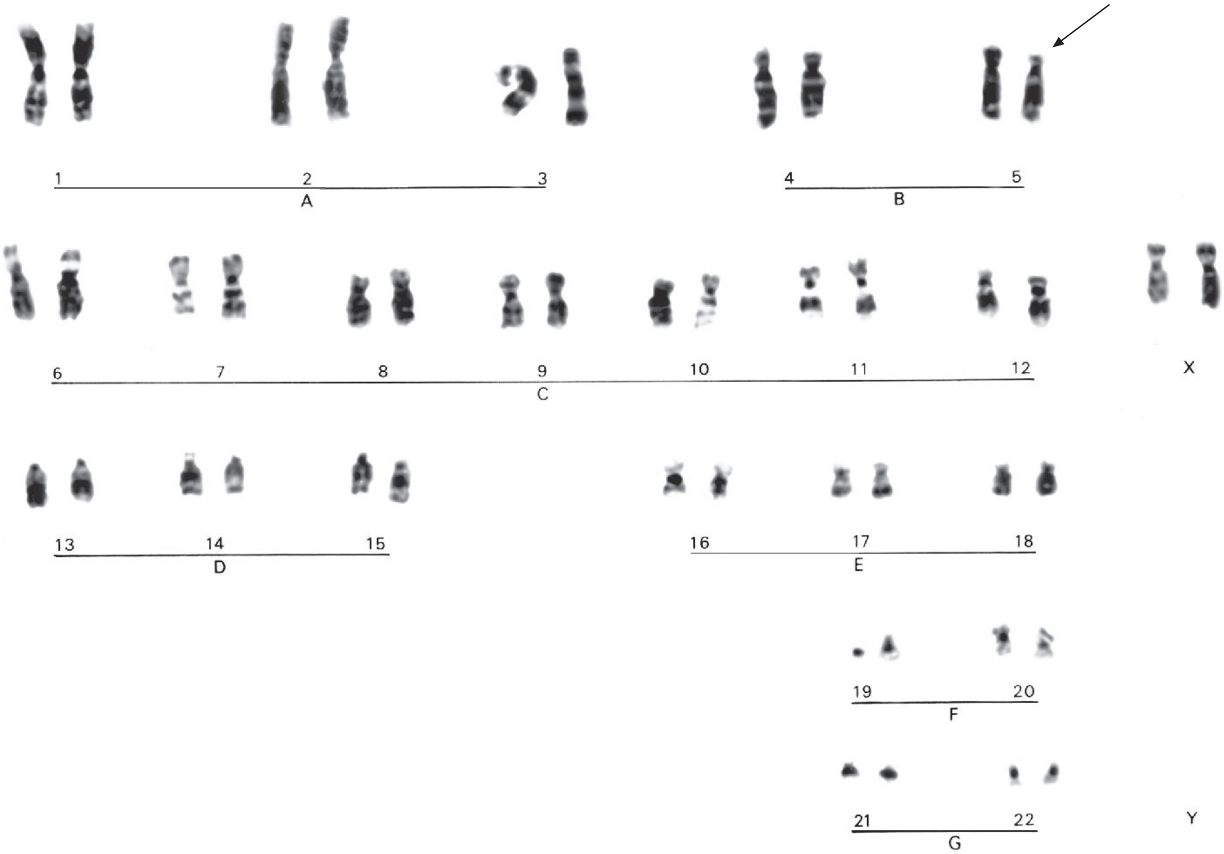

1
2
3
A
4
5
B
6
7
8
9
10
11
12
C
X
13
14
15
D
16
17
18
E
19
20
F
21
22
G
Y

**Figure 8.16** Karyotype of individual with cri du chat syndrome, due to a partial deletion of the short arm of chromosome 5 (5p−; *arrow*). (Reproduced courtesy of Dr. Thomas G. Brewster, Foundation for Blood Research, Scarborough, Maine.)

other deletions studied (4p–, 13q–, 18p–, 18q–) also result in microcephaly and severe mental retardation. The rarity of viable deletion heterozygotes is consistent with the fact that viable monosomics (having a single chromosome of a pair) are rare. An individual heterozygous for a deletion is, in effect, monosomic for the deleted region of the chromosome. Evidently, monosomy or heterozygosity for larger deleted regions of a chromosome is generally lethal in human beings.

## VARIATION IN CHROMOSOME NUMBER

Anomalies of chromosome number occur as either **euploidy** or **aneuploidy.** Euploidy involves changes in whole *sets* of chromosomes; aneuploidy involves changes in chromosome *number* by additions or deletions of less than a whole set.

### Aneuploidy

An explanation for the terminology of aneuploid change appears in table 8.1. A diploid cell missing a single chromosome is **monosomic.** A cell missing both copies of that chromosome is **nullisomic.** A cell missing two nonhomologous chromosomes is a double monosomic. A similar terminology exists for extra chromosomes. For example, a diploid cell with an extra chromosome is **trisomic.** Aneuploidy results from nondisjunction in meiosis or by chromosomal lagging whereby one chromosome moves more slowly than the others during anaphase, is excluded from the telophase nucleus, and is thus lost. Here, nondisjunction is illustrated using the sex chromosomes in XY organisms such as human beings or fruit flies. Four examples are shown (fig. 8.17): nondisjunction in either the male or female at either the first or second meiotic divisions. Figure 8.18 shows the types of zygotes that can result when these nondisjunctional gametes fuse with normal gametes. All of the offspring produced are chromosomally abnormal. The names and kinds of these imbalances in human beings are detailed later in this chapter.

Bridges first showed the occurrence of nondisjunction in *Drosophila* in 1916 with crosses involving the white-eye locus. When a white-eyed female was crossed with a wild-type male, typically the daughters were wild-type and the sons were white-eyed. However, occasionally (one or two per thousand), a white-eyed daughter or a wild-type son appeared. This could be explained most easily by a nondisjunctional event in the white-eyed females, where $X^wX^w$ and 0 eggs (without sex chromosomes) were formed. Under this hypothesis, if a Y-bearing sperm fertilized an $X^wX^w$ egg, the offspring would be an $X^wX^wY$ white-eyed daughter. If a normal $X^+$-bearing sperm fertilized the egg without sex chromosomes, the result would be an $X^+0$ wild-type son. Subsequently, these exceptional individuals were found by cytological examination to have precisely the predicted chromosomes (XXY daughters and X0 sons). The other types produced by this nondisjunctional event are the XX egg fertilized by an X-bearing sperm and the 0 egg fertilized by the Y-bearing sperm. The XXX zygotes are genotypically $X^wX^wX^+$, or wild-type daughters (which usually die), and Y0 flies (which always die).

### Mosaicism

Rarely, an individual is made up of several cell lines, each with different chromosome numbers. These individuals are referred to as **mosaics** or **chimeras,** depending on the sources of the cell lines. Such conditions can be the result of nondisjunction or chromosomal lagging during mitosis in the zygote or in nuclei in the early embryo (mosaic). This is demonstrated, again for sex chromosomes, in figure 8.19. A lagging chromosome is shown in figure 8.20; in the figure, the X chromosome is lost in one

**Table 8.1** **Partial List of Terms to Describe Aneuploidy, Using *Drosophila* as an Example (Eight Chromosomes: X, X, 2, 2, 3, 3, 4, 4)**

| Type | Formula | Number of Chromosomes | Example |
|---|---|---|---|
| Normal | $2n$ | 8 | X, X, 2, 2, 3, 3, 4, 4 |
| Monosomic | $2n - 1$ | 7 | X, X, 2, 2, 3, 4, 4 |
| Nullisomic | $2n - 2$ | 6 | X, X, 2, 2, 4, 4 |
| Double monosomic | $2n - 1 - 1$ | 6 | X, X, 2, 3, 4, 4 |
| Trisomic | $2n + 1$ | 9 | X, X, 2, 2, 3, 3, 4, 4, 4 |
| Tetrasomic | $2n + 2$ | 10 | X, X, 2, 2, 3, 3, 3, 3, 4, 4 |
| Double trisomic | $2n + 1 + 1$ | 10 | X, X, 2, 2, 2, 3, 3, 3, 4, 4 |

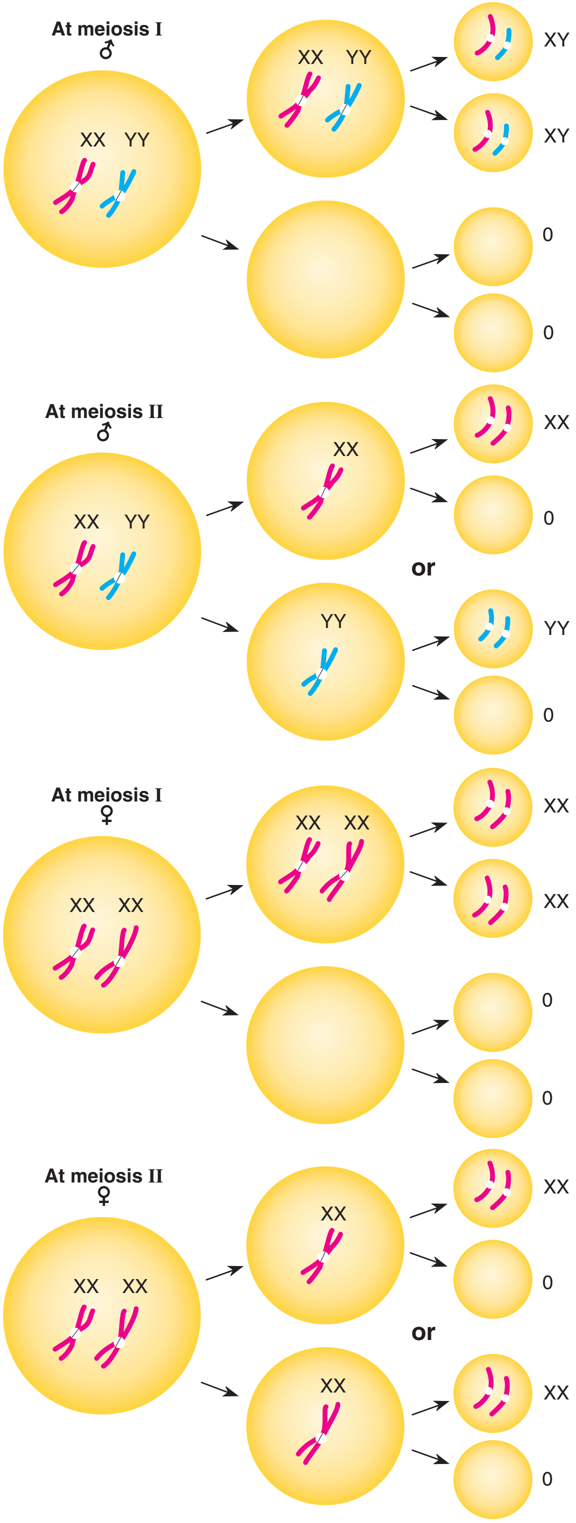

At meiosis I
♂
XX YY
XX YY
XY
XY
0
0
At meiosis II
♂
XX YY
XX
XX
0
or
YY
YY
0
At meiosis I
♀
XX XX
XX XX
XX
XX
0
0
At meiosis II
♀
XX XX
XX
XX
0
or
XX
XX
0

**Figure 8.17** Nondisjunction of the sex chromosomes in *Drosophila* or human beings. "0" refers to the lack of sex chromosomes.

of the dividing somatic cells, resulting in an XX cell line and an X0 cell line. In *Drosophila,* if this chromosomal lagging occurs early in development, an organism that is part male (X0) and part female (XX) develops. Figure 8.21 shows a fruit fly in which chromosomal lagging has occurred at the one-cell stage, causing the fly to be half male and half female. A mosaic of this type, involving male and female phenotypes, has a special name—**gynandromorph.** (A **hermaphrodite** is an individual, not necessarily mosaic, with both male and female reproductive organs.) Many sex-chromosomal mosaics are known in humans, including XX/X, XY/X, XX/XY, and XXX/X. At least one case is known of a human XX/XY chimera that resulted from the fusion of two zygotes, one

| | | Nondisjunction | | | | | |
|---|---|---|---|---|---|---|---|
| | | ♂ | | | | ♀ | |
| | | XY | XX | YY | 0 | XX | 0 |
| Normal | ♀X | XXY | XXX | XYY | X0 | | |
| | ♂X | | | | | XXX | X0 |
| | ♂Y | | | | | XXY | Y0 |

**Figure 8.18** Results of fusion of a nondisjunction gamete (*top*) with a normal gamete (*side*).

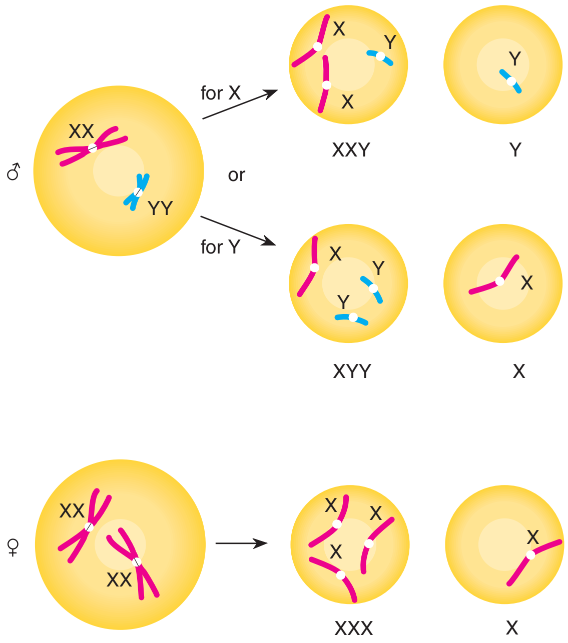

♂
XX
YY
for X
or
for Y
X
Y
X
XXY
Y
Y
X
Y
Y
XYY
X
X
♀
XX
XX
X
X
X
XXX
X
X

**Figure 8.19** Mitotic nondisjunction of the sex chromosomes.

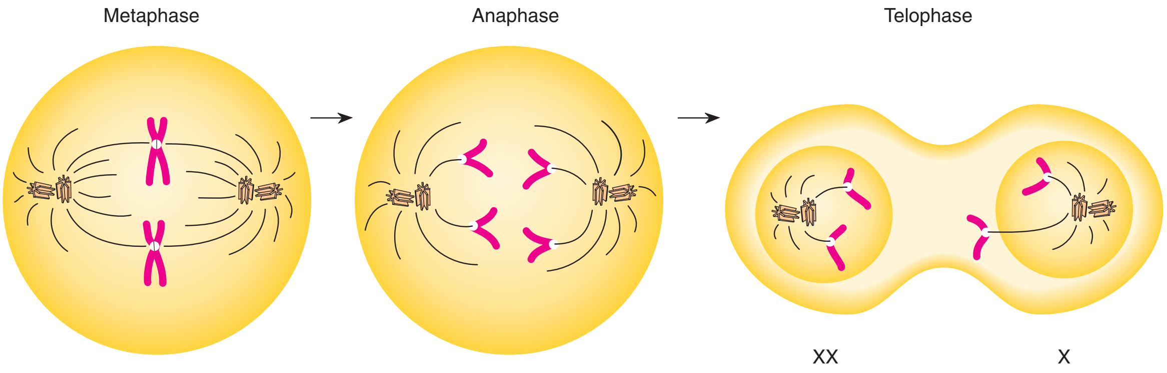


**Figure 8.20** Chromosomal lagging at mitosis in the X chromosomes of a female *Drosophila.*

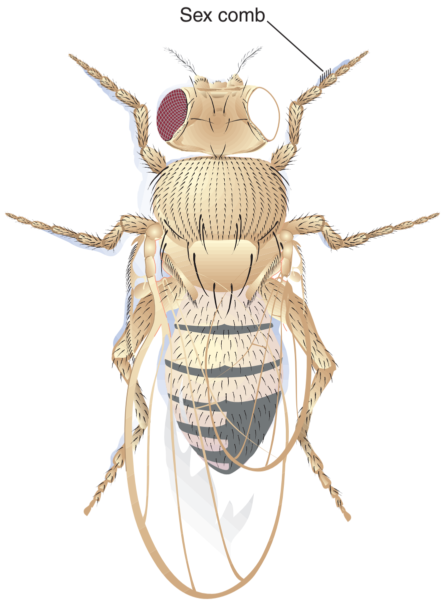


**Figure 8.21** *Drosophila* gynandromorph. The *left side* is wild-type XX female; the *right side* is X0 male, hemizygous for white eye and miniature wing.

formed by a sperm fertilizing an ovum and the other formed by a second sperm fertilizing a polar body of that ovum.

## Aneuploidy in Human Beings

Approximately 50% of spontaneous abortions (miscarriages) among women in the United States involve fetuses with some chromosomal abnormality; about half of these are autosomal trisomics. About one in 160 live human births has some sort of chromosomal anomaly; most are balanced translocations, autosomal trisomics, or sex-chromosomal aneuploids.

In the standard system of nomenclature, a normal human chromosome complement is 46,XX for a female and 46,XY for a male. The total chromosome number appears first, then the description of the sex chromosomes, and, finally, a description of autosomes if some autosomal anomaly is evident. For example, a male with an extra X chromosome would be 47,XXY. A female with a single X chromosome would be 45,X. Since all the autosomes are numbered, we describe their changes by referring to their addition (+) or deletion (−). For example, a female with trisomy 21 would be 47,XX,+21. The short arm of a chromosome is designated p, the longer arm, q. When a change in part of the chromosome occurs, a + after the arm indicates an increase in the length of that arm, whereas a minus sign (−) indicates a decrease in its length. For example, a translocation (t) that transfers part of the short arm of chromosome 9 to the short arm of chromosome 18 would be 46,XX, t(9p−;18p+). The semicolon indicates that both chromosomes kept their centromeres.

Following are descriptions of viable human aneuploids who survive long enough after birth to have a named syndrome.

### *Trisomy 21 (Down Syndrome), 47,XX or XY,+21*

Down syndrome (figs. 8.22 and 8.23) affects about one in seven hundred live births. Most affected individuals are mildly to moderately mentally retarded and have congenital heart defects and a very high (1/100) risk of acute leukemia. They are usually short and have a broad, short skull; hyperflexibility of joints; and excess skin on the back of the neck. The physician John Langdon Down first described this syndrome in 1866. (Modern convention is to avoid the possessive form of a name in referring to a syndrome.) Down syndrome was the first human syndrome attributed to a chromosomal disorder; Jérôme Lejeune, a physician in Paris, published this finding in

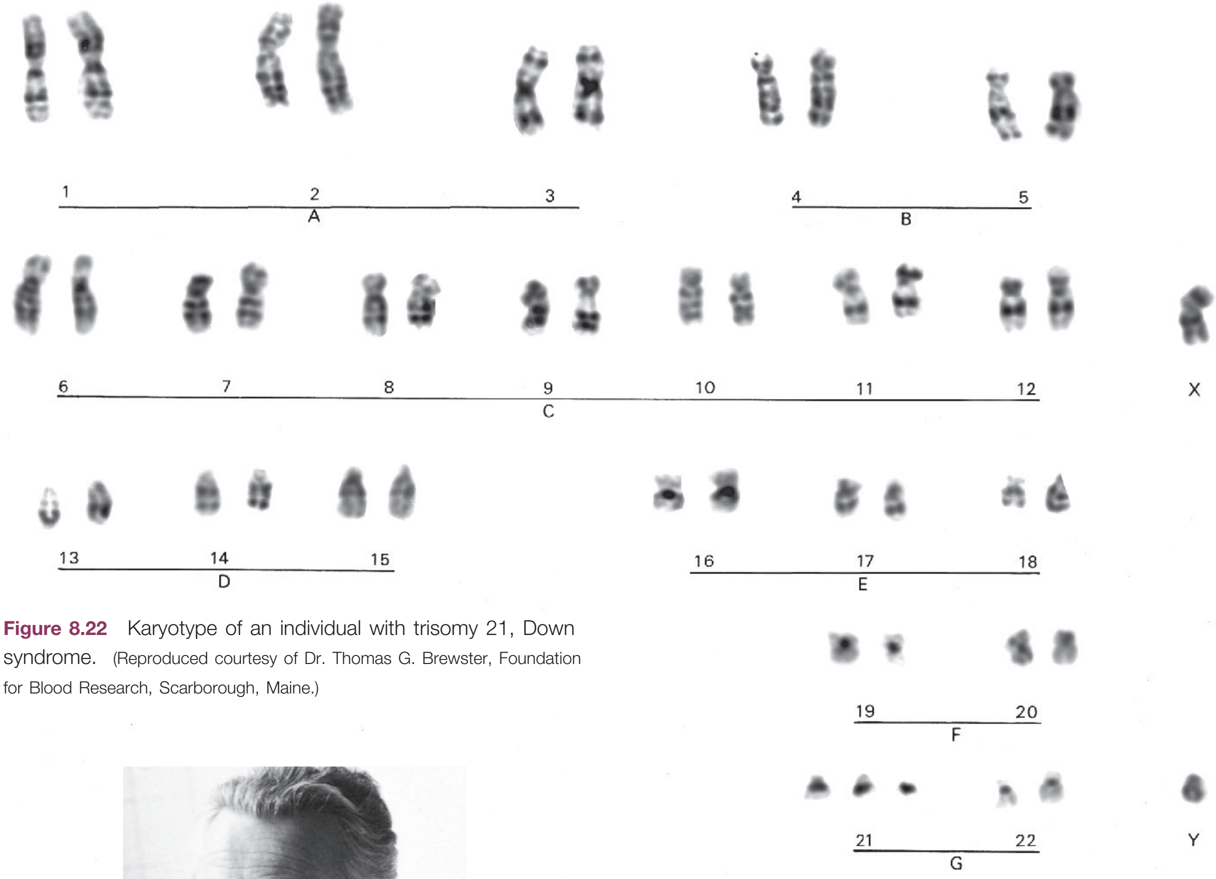


**Figure 8.22** Karyotype of an individual with trisomy 21, Down syndrome. (Reproduced courtesy of Dr. Thomas G. Brewster, Foundation for Blood Research, Scarborough, Maine.)

Jérôme Lejeune (1926–1994).
(Courtesy of Dr. Jérôme Lejeune, Institut de Progenese, Paris.)

1959. An interesting aspect of this syndrome is its increased incidence among children of older mothers (fig. 8.24), a fact known more than twenty-five years before the discovery of the cause of the syndrome. Since the future ova are in prophase I of meiosis (dictyotene) since before the mother's birth, all ova are the same age as the female. Presumably, older ova are more susceptible to nondisjunction of chromosome 21.

Recently, techniques of molecular genetics (chapter 13) have been used to identify the origins of the three copies of chromosome 21 in a large sample of individuals with Down syndrome. As expected, the overwhelming majority of the extra copies of chromosome 21 (95%) were of maternal origin. About 5% of the cases of Down syndrome were of mitotic origin, occurring either in the gonad of one of the parents (evenly split between mothers and fathers) or possibly postzygotically in the fetus.

### *Familial Down Syndrome*

Down syndrome (trisomy 21), as described, is usually the result of either a nondisjunctional event during gametogenesis or, rarely, a mitotic event. It is a function of maternal age and is not inherited. (Although about half the children of a person with trisomy 21 will have trisomy 21 because of aneuploid gamete production, the possibility that an unaffected relative of the person will have abnormal children is no greater than for a person of the same age chosen at random from the general population.) However, about 4% of those with Down syndrome have been found to have a translocation of chromosome 21,

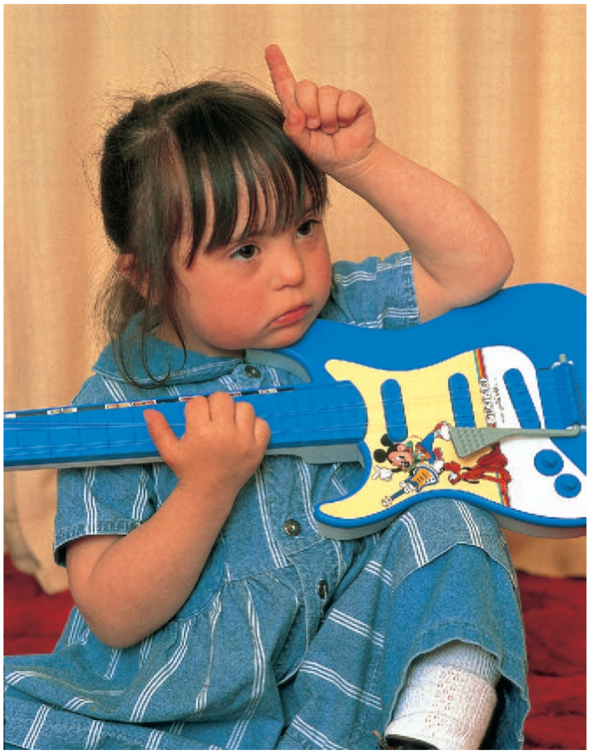
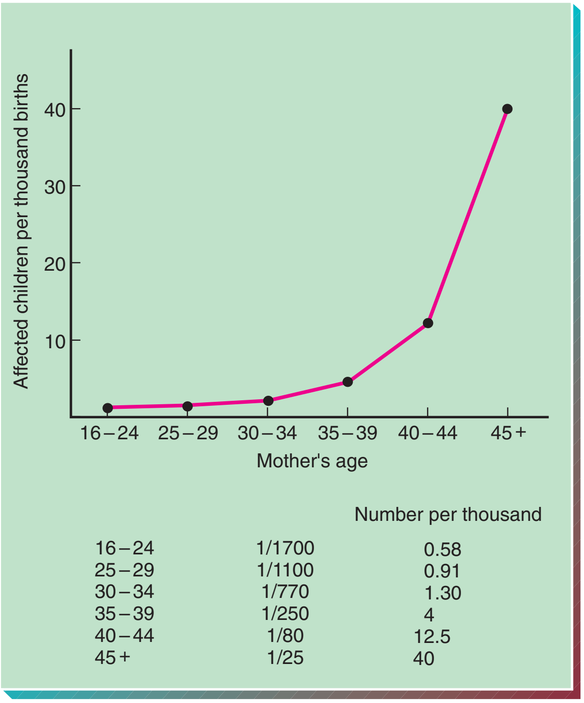


**Figure 8.23** Individual with trisomy 21. (© Hattie Young/SPL/Photo Researchers.)



| | | Number per thousand |
|---|---|---|
| 16–24 | 1/1700 | 0.58 |
| 25–29 | 1/1100 | 0.91 |
| 30–34 | 1/770 | 1.30 |
| 35–39 | 1/250 | 4 |
| 40–44 | 1/80 | 12.5 |
| 45+ | 1/25 | 40 |

**Figure 8.24** Increased risk of trisomy 21 attributed to the age of the mother. (From E. Hook, "Estimates of Maternal Age-Specific Risks of a Down-Syndrome Birth in Women age 34–41," *Lancet,* 2:33–34, Copyright © 1976 by The Lancet Ltd.)

usually associated with chromosome 14, 15, or 22. The translocational and nontranslocational types of Down syndrome have identical symptoms; however, a balanced translocation can be passed on to offspring (see fig. 8.11). Alternate segregation of centromeres in the translocation heterozygote produces either a normal gamete or one carrying the balanced translocation. Adjacent segregation causes partial trisomy for certain chromosomal parts. When this occurs for most of chromosome 21, Down syndrome results.

It is worth mentioning that aside from trisomy and translocation, Down syndrome can come about through mosaicism, as mentioned earlier, or a centromeric event. About 2% of individuals with Down syndrome are mosaic for cells with both two and three copies of chromosome 21. Some evidence suggests that the original zygotes were trisomic, but then a daughter cell lost one of the copies of chromosome 21. The severity of the symptoms in these individuals relates to the percentage of trisomic cells they possess. Mosaicism increases with maternal age, just as trisomy in general does. In extremely rare cases, Down syndrome is caused by an abnormal chromosome 21 that has, rather than a short and long arm, two identical long arms attached to the centromere. This type of chromosome, called an **isochromosome,** presumably occurs by an odd centromeric fission (fig. 8.25). Hence, a person with a normal chromosome 21 and an isochromosome 21 has three copies of the long arm of the chromosome and has Down syndrome.

### *Trisomy 18 (Edward Syndrome), 47,XX or XY,+18*

Edward syndrome affects one in ten thousand live births (fig. 8.26). Most affected individuals are female, with 80 to 90% mortality by two years of age. The infant usually has an elfin appearance with small nose and mouth, a receding lower jaw, abnormal ears, and a lack of distal flexion creases on the fingers. The distal joints have limited motion, and the fingers display a characteristic posturing in which the little and index fingers overlap the middle two. The syndrome is usually accompanied by severe mental retardation.

### *Trisomy 13 (Patau Syndrome), 47,XX or XY,+13, and Other Trisomic Disorders*

Patau syndrome affects one in twenty thousand live births. Diagnostic features are cleft palate, cleft lip, con-

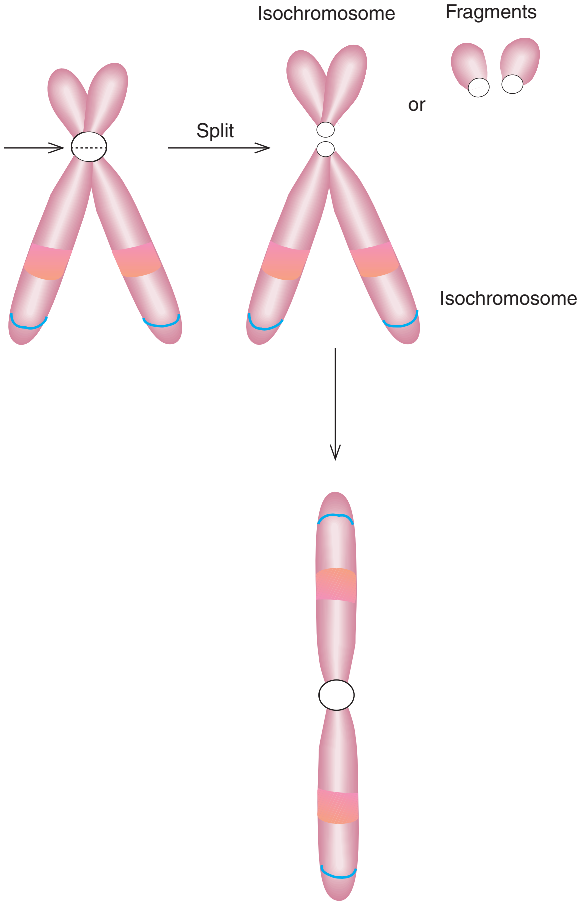


**Figure 8.25** If the centromere of chromosome 21 breaks perpendicular to the normal division axis, it can form an isochromosome of the long arms and either an isochromosome of the short arms or two separate fragments. This can happen during anaphase of mitosis or meiosis II.

genital heart defects, polydactyly, and severe mental retardation. Mortality is very high in the first year of life.

Other autosomal trisomics are known but are extremely rare. These include trisomy 8 (47,XX or XY,+8) and cat's eye syndrome, a trisomy of an unidentified, small acrocentric chromosome (47,XX or XY,[+acrocentric]). Several aneuploids involving sex chromosomes are also known.

### *Turner Syndrome, 45,X*

About one in ten thousand live female births is of an infant with Turner syndrome. This and 45,XX or XY,−21 and 45,XX or XY,−22 are the only nonmosaic, viable monosomics recorded in human beings (fig. 8.27), indicating the severe consequences monosomy has on all but the two smallest autosomes and a sex chromosome. Individuals with Turner syndrome usually have normal intelligence but underdeveloped ovaries, abnormal jaws, webbed necks, and shieldlike chests.

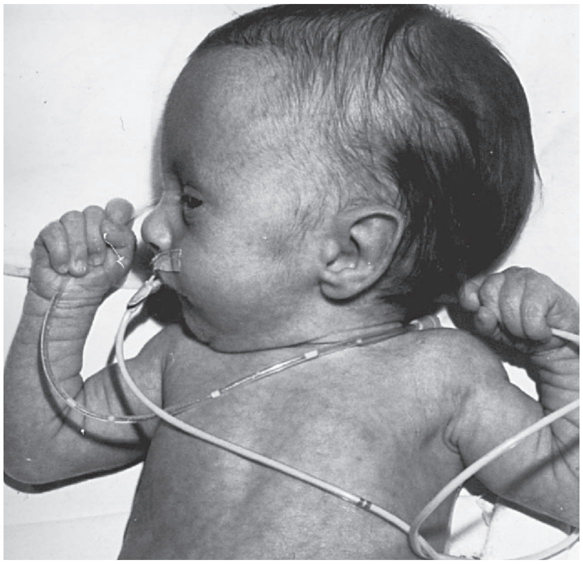


**Figure 8.26** Child with trisomy 18, Edward syndrome.
(Reproduced courtesy of Dr. Jérôme Lejeune, Institut de Progenese, Paris.)

The symptoms of Turner syndrome have been logically deduced to be caused by a single dosage of genes that are normally present and active in two dosages. Thus, these genes would be located on both the X and Y chromosomes (pseudoautosomal) to provide two dosages in normal XY males and also be active in both X chromosomes in normal XX females. Therefore, they should be located on regions of the X chromosome that escape inactivation (see chapter 5). Studies of persons with small X-chromosomal deletions and molecular analyses of the X and Y chromosomes (outlined in chapter 13) have caused two genes to emerge as candidates: *ZFY* (on the Y chromosome, termed *ZFX* on the X chromosome) and *RPS4Y* (on the Y chromosome, termed *RPS4X* on the X chromosome). *ZFY* (zinc finger on the Y chromosome) was once believed to be the male-determining gene in mammals. *RPS4Y* encodes a ribosomal protein, one of the many proteins making up the ribosome.

It is interesting to note a dosage-compensation difference in people and mice, which have analogous genes termed *Zfx* and *Rps4x.* In mice, unlike in people, these genes are inactivated in the "Lyonized" X chromosome in females and have restricted activity in the Y

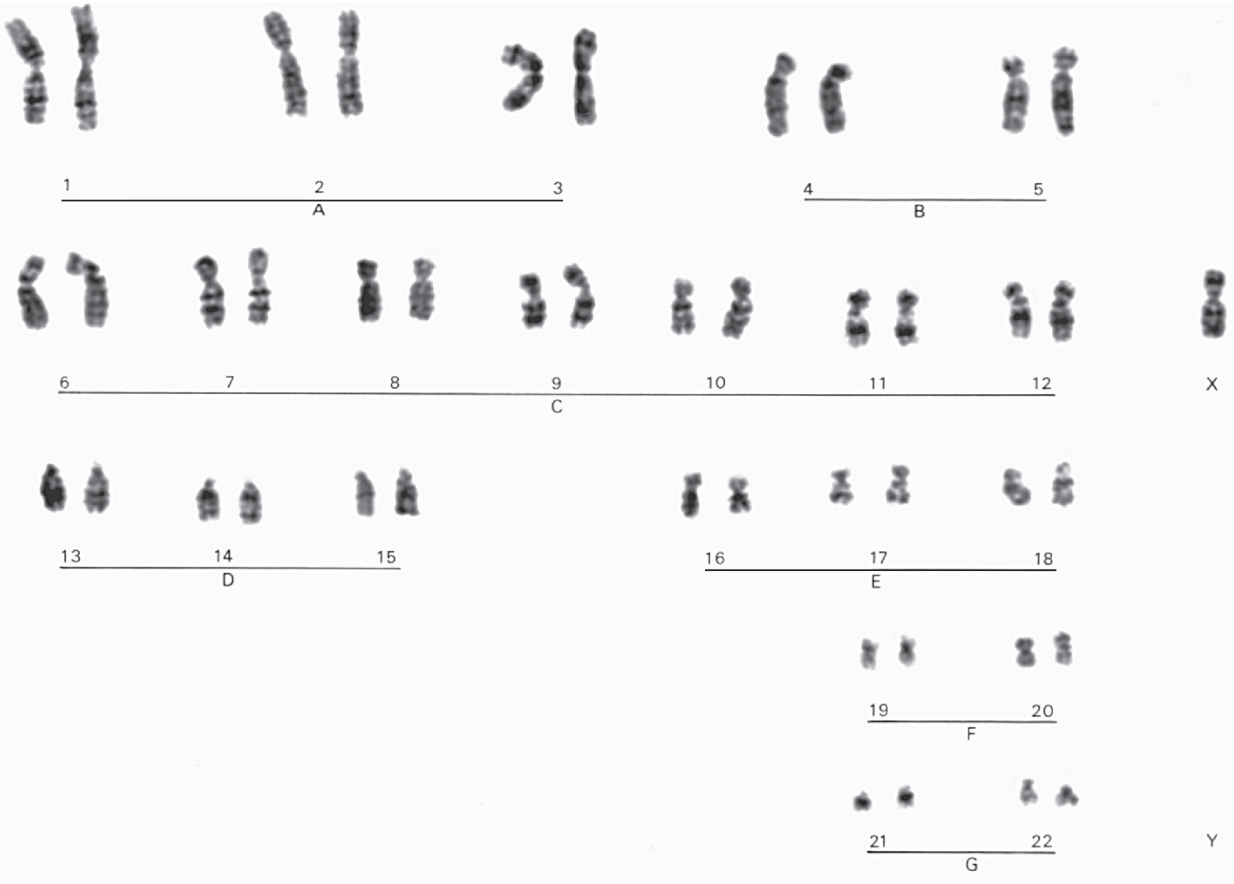


**Figure 8.27** Karyotype of a person with Turner syndrome (X0). (Reproduced courtesy of Dr. Thomas G. Brewster, Foundation for Blood Research, Scarborough, Maine.)

chromosome. Hence, mouse cells seem normally to have only one copy of these genes functioning in normal XY males and XX females. Therefore, we would predict that the X0 genotype in mice would produce few if any negative effects as compared with a human X0 genotype, since mouse cells of both sexes normally only have one functional copy of each gene. In fact, human Turner syndrome fetuses have a 99% prenatal mortality rate, but virtually no prenatal mortality affects mouse fetuses with the X0 genotype (born of XX mothers). This confirms our predictions and points to differences between people and mice in dosage-compensation mechanisms for specific genes.

### *XYY Karyotype, 47,XYY*

About one in one thousand live male births is of an individual with an XYY karyotype. (We avoid the term *syndrome* here because XYY men have no clearly defined series of attributes, other than often being taller than normal.) Some controversy has surrounded this karyotype because it was once reported that it occurred in abundance in a group of mentally subnormal males in a prison hospital. Seven XYY males were found among 197 inmates, whereas only one in about two thousand control men were XYY. This study has subsequently been expanded and corroborated. Although it is now fairly well established that the incidence of XYY males in prison is about twentyfold higher than in society at large, the statistic is somewhat misleading: the overwhelming number of XYY men seem to lead normal lives. At most, about 4% of XYY men end up in penal or mental institutions, where they make up about 2% of the population.

There is some indication that the XYY men in prison had lower intelligence test levels. Thus, criminal tendency may be attributed to lower intelligence rather than a predisposition toward criminality caused by an extra Y chromosome. For the most part, expanded studies have indicated that XYY criminals do not commit violent crimes.

Research on this karyotype has produced its own problems. A research project at Harvard University on XYY males came under intense public pressure and was eventually terminated. The project, under the direction of Stanley Walzer (a psychiatrist) and Park Gerald (a geneticist), involved screening all newborn boys at the Boston Hospital for Women and following the development of those with chromosomal anomalies. The criticism of this work centered mainly on the necessity of informing parents that their sons had an XYY karyotype that might be associated with behavioral problems. Opponents of this work claimed that telling the parents could trigger a self-fulfilling prophecy; that is, parents who heard that their children were not normal and *might* cause trouble might then behave toward their children in a manner that would increase the probability that their children *would* cause trouble. The opponents claimed that the risks of this research outweighed the benefits. The project was terminated in 1975 primarily because of the harassment Walzer faced.

### Klinefelter Syndrome, 47,XXY

The incidence of Klinefelter syndrome is about one in one thousand live births. Tall stature and infertility are common symptoms. Diagnosis is usually by buccal (cheek tissue) smear to ascertain the presence of a Barr body in a male, indicating an XXY karyotype. Some problems with behavior and speech development are associated with this syndrome.

### Triple-X Female, 47,XXX, and Other Aneuploid Disorders of Sex Chromosomes

A triple-X female appears in about one in one thousand female live births. Fertility can be normal, but these individuals are usually mildly mentally retarded. Delayed growth, as well as congenital malformations, are also sometimes present. Other sex-chromosomal aneuploids, including XXXX, XXXXX, and XXXXY, are extremely rare. All seem to be characterized by mental retardation and growth deficiencies.

## Euploidy

Euploid organisms have varying numbers of complete haploid chromosomal sets. We are already familiar with haploids ($n$) and diploids ($2n$). Organisms with higher numbers of sets, such as **triploids** ($3n$) and **tetraploids** ($4n$), are called polyploids. Three kinds of problems plague polyploids. First, the potential exists for a general imbalance in the organism due to the extra genetic material in each cell. For example, a triploid human fetus has about a one in a million chance to survive to birth, at which time death usually occurs due to problems in all organ systems. Second, if there is a chromosomal sex-determining mechanism, it may be disrupted by polyploidy. And third, meiosis produces unbalanced gametes in many polyploids.

If the polyploid has an odd number of sets of chromosomes, such as triploid ($3n$), two of the three homologues will tend to pair at prophase I of meiosis, producing a bivalent and a univalent. The bivalent separates normally, but the third chromosome goes independently to one of the poles. This separation results in a 50% chance of aneuploidy in each of the $n$-different chromosomes, rapidly decreasing the probability of a balanced gamete as $n$ increases. Therefore, as $n$ increases, so does the likelihood of sterility. An alternative to the bivalent-univalent type of synapsis is the formation of trivalents, which have similar problems (fig. 8.28). Even-numbered polyploids, such as tetraploids ($4n$), can do better during meiosis. If the centromeres segregate two by two in each of the $n$ meiotic figures, balanced gametes can result. Often, however, the multiple copies of the chromosomes form complex figures during synapsis, including monovalents, bivalents, trivalents, and quadrivalents, tending to result in aneuploid gametes and sterility.

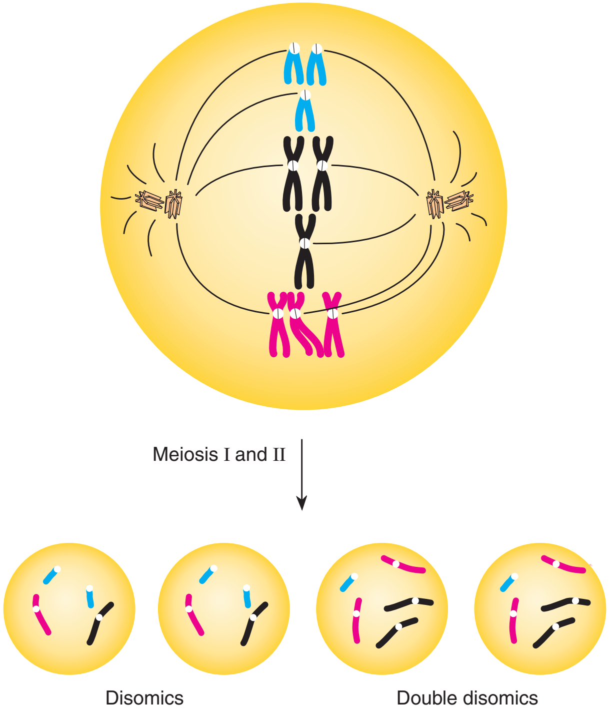


**Figure 8.28** Meiosis in a triploid ($3n = 9$) and one possible resulting arrangement of gametes. The probability of a "normal" gamete is $(1/2)^n$ where $n$ equals the haploid chromosome number. Here, $n = 3$ and $(1/2)^3 = 1/8$.

Some groups of organisms, primarily plants, have many polyploid members. An estimated 30 to 80% of all flowering plant species (angiosperms) are polyploids, as are 95% of ferns. (Polyploidy is apparently rare in gymnosperms and fungi.) For example, the genus of wheat, *Triticum,* has members with fourteen, twenty-eight, and forty-two chromosomes. Because the basic *Triticum* chromosome number is $n = 7$, these forms are $2n$, $4n$, and $6n$ species, respectively. Chrysanthemums have species of eighteen, thirty-six, fifty-four, seventy-two, and ninety chromosomes. With a basic number of $n = 9$, these species represent a $2n$, $4n$, $6n$, $8n$, and $10n$ series. In both these examples, the even-numbered polyploids are viable and fertile, but the odd-numbered polyploids are not.

### Autopolyploidy

Polyploidy can come about in two different ways. In **autopolyploidy,** all of the chromosomes come from within the same species. In **allopolyploidy,** the chromosomes come from the hybridization of two different species (fig. 8.29). Autopolyploidy occurs in several different ways. The fusion of nonreduced gametes creates polyploidy. For example, if a diploid gamete fertilizes a normal haploid gamete, the result is a triploid. Similarly, if a diploid gamete fertilizes another diploid gamete, the result is a tetraploid. The equivalent of a nonreduced gamete comes about in meiosis if the parent cell is polyploid to begin with. For example, if one branch of a diploid plant is tetraploid, its flowers produce diploid gametes. These gametes are not the result of a failure to reduce chromosome numbers meiotically, but rather the result of successful meiotic reduction in a polyploid flower. The tetraploid tissue of the plant in this example can originate by the **somatic doubling** of diploid tissues.

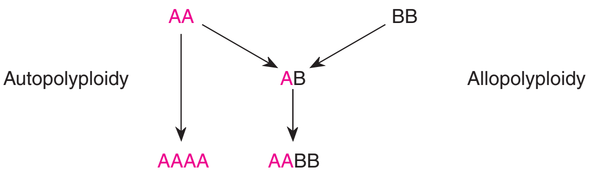


**Figure 8.29** Autopolyploidy and allopolyploidy. If A and B are the haploid genomes of species A and B, respectively, then autopolyploidy produces a species with an AAAA karyotype, and allopolyploidy (with chromosome doubling) produces a species with AABB karyotype. If A represents seven chromosomes, then an AA diploid has fourteen chromosomes and an AAAA tetraploid has twenty-eight chromosomes. If B represents five chromosomes, then a BB diploid has ten chromosomes and an AABB allotetraploid has twenty-four chromosomes.

Somatic doubling can come about spontaneously or be caused by anything that disrupts the normal sequence of a nuclear division. For example, colchicine induces somatic doubling by inhibiting microtubule formation. This prevents the formation of a spindle and thus prevents the chromosomes from moving apart during either mitosis or meiosis. The result is a cell with double the chromosome number. Other chemicals, temperature shock, and physical shock can produce the same effect.

### Allopolyploidy

Allopolyploidy comes about by cross-fertilization between two species. The resulting offspring have the sum of the reduced chromosome number of each parent species. If each chromosome set is distinctly different, the new organisms have difficulty in meiosis because no two chromosomes are sufficiently homologous to pair. Then every chromosome forms a univalent (unpaired) figure, and they separate independently during meiosis, producing aneuploid gametes. However, if an organism can survive by vegetative growth until somatic doubling takes place in gamete precursor cells ($2n \rightarrow 4n$), or alternatively, if the zygote was formed by two unreduced gametes ($2n + 2n$), the resulting offspring will be fully fertile because each chromosome has a pairing partner at meiosis. We can draw an example from the work of Russian geneticist G. D. Karpechenko.

In 1928, Karpechenko worked with the radish (*Raphanus sativus,* $2n = 18, n = 9$) and cabbage (*Brassica oleracea,* $2n = 18, n = 9$). When these two plants are crossed, an $F_1$ results with $n + n = 18$ (9 + 9). This plant, which is an allodiploid, has characteristics intermediate between the two parental species (fig. 8.30). If somatic doubling takes place, the chromosome number is doubled to thirty-six, and the plant becomes an allopolyploid (an allotetraploid of $4n$). Since each chromosome has a homologue, this allotetraploid is also referred to as an **amphidiploid.** If we did not know its past history, this plant would simply be classified as a diploid with $2n = 36$. In this case, the new amphidiploid cannot successfully breed with either parent because the offspring are sterile triploids. It is, therefore, a new species and has been named *Raphanobrassica.* As an agricultural experiment, however, it was not a success because it did not combine the best features of the cabbage and radish.

### Polyploidy in Plants and Animals

Although polyploids in the animal kingdom are known (in some species of lizards, fish, invertebrates, and a

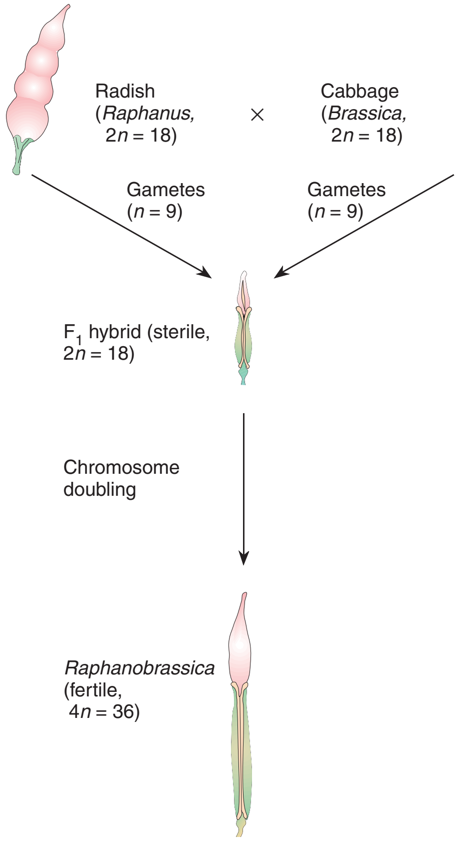


**Figure 8.30** Hybridization of cabbage and radish, showing the resulting hybrid fruiting structures.

tetraploid mammal, the red viscacha rat), polyploidy as a successful evolutionary strategy is primarily a plant phenomenon. There are several reasons for this. To begin with, many more animals than plants have chromosomal sex-determining mechanisms. Polyploidy severely disrupts these mechanisms. For example, Bridges discovered a tetraploid female fruit fly, but it has not been possible to produce a tetraploid male. The tetraploid female's progeny were triploids and intersexes. A second reason why polyploidy is more common in plants is because plants can generally avoid the meiotic problems of polyploidy longer than most animals. Some plants can exist vegetatively, allowing more time for the rare somatic doubling event to occur that will produce an amphidiploid; animal life spans are more precisely defined, allowing less time for a somatic doubling. And third, many plants depend on the wind or insect pollinators to fertilize them and thus have more of an opportunity for hybridization. Many animals have relatively elaborate courting rituals that tend to restrict hybridization.

Polyploidy has been used in agriculture to produce "seedless" as well as "jumbo" varieties of crops. Seedless watermelon, for example, is a triploid. Its seeds are mostly sterile and do not develop. It is produced by growing seeds from the cross between a tetraploid variety and a diploid variety. Jumbo Macintosh apples are tetraploid.

# SUMMARY

**STUDY OBJECTIVE 1:** To observe the nature and consequences of chromosomal breakage and reunion 178–190

Variation can occur in the structure and number of chromosomes in the cells of an organism. When chromosomes break, the ends become "sticky"; they tend to reunite with other broken ends. A single break can lead to deletions or the formation of acentric or dicentric chromosomes. Dicentrics tend to go through breakage-fusion-bridge cycles, which result in duplications and deficiencies.

Two breaks in the same chromosome can yield deletions and inversions. Variegation position effects, as well as new linkage arrangements, can result. Inversion heterozygotes produce loop figures during synapsis, which can form either at meiosis or in polytene chromosomes. Heterozygosity for an inversion suppresses crossovers; organisms that are heterozygotes are semisterile.

Reciprocal translocations can result from single breaks in nonhomologous chromosomes. These produce cross-shaped figures at synapsis and result in semisterility. The *Bar* eye phenotype of *Drosophila* is an example of a duplication that causes a position effect.

**STUDY OBJECTIVE 2:** To observe the nature and consequences of variation in chromosome numbers in human and nonhuman organisms 190–199

Changes in chromosome number can involve whole sets (euploidy) or partial sets (aneuploidy) of chromosomes. Aneuploidy usually results from nondisjunction or chromosomal lagging. Several medical syndromes, such as Down, Turner, and Klinefelter syndromes, and the XYY karyotype are caused by aneuploidy.

Polyploidy leads to difficulties in chromosomal sex-determining mechanisms, general chromosomal imbalance, and problems during meiotic segregation. It has been more successful in plants than in animals because plants generally lack chromosomal sex-determining mechanisms. Plants can also avoid meiotic problems by propagating vegetatively. In both animals and plants, even-numbered polyploids do better than odd-numbered polyploids because they have a better chance of producing balanced gametes during meiosis. Somatic doubling provides each chromosome in a hybrid organism with a homologue, and thus makes possible tetrad formation at meiosis. New species have arisen by polyploidy.

## SOLVED PROBLEMS

**PROBLEM 1:** What are the consequences of an inversion?

***Answer:*** In an inversion homozygote, the consequences are change in linkage arrangements, including new orders and map distances, and the possibility of position effects if a locus is placed into or near heterochromatin. In an inversion heterozygote, crossover suppression causes semisterility because zygotes that carry genic imbalances are lost. Inversion heterozygotes can be seen as meiotic loop structures or loops formed in endomitotic chromosomes such as those found in the salivary glands of fruit flies. In an evolutionary sense, inversions result in supergenes, locking together allelic combinations.

**PROBLEM 2:** What are the consequences of a monosomic chromosome in human beings?

***Answer:*** In human beings, monosomy is rare, meaning that, with few exceptions, it is lethal. In fact, monosomics are also rare in spontaneous abortions, indicating that most monosomic fetuses are lost before the woman is aware of the pregnancy. The only monosomics known to be viable in human beings are Turner syndrome (45,X) and monosomics of chromosomes 21 and 22, the two smallest autosomal chromosomes.

**PROBLEM 3:** Ebony body (*e*) in flies is an autosomal recessive trait. A true-breeding ebony female (*ee*) is mated with a true-breeding wild-type male that has been irradiated. Among the wild-type progeny is a single ebony male. Explain this observation.

***Answer:*** The cross is $ee \times e^+e^+$, and all $F_1$s should be $e^+e$ (wild-type). The use of irradiation alerts us to the possibility of chromosomal breaks, as well as simple mutations. What type of chromosomal aberration would allow a recessive trait to appear unexpectedly? A deletion, which creates pseudodominance when there is no second allele, is a good possibility. The male in question could have gotten the ebony allele from its mother and no homologous allele from its father. Alternatively, the wild-type allele from the father could have mutated to an ebony allele.

## EXERCISES AND PROBLEMS*

**VARIATION IN CHROMOSOMAL STRUCTURE**

1. What kind of figure is observed in meiosis of a reciprocal translocation homozygote?
2. Can a deletion result in the formation of a variegation position effect? If so, how?
3. Does crossover suppression occur in an inversion homozygote? Explain.
4. Which rearrangements of chromosomal structure cause semisterility?
5. What are the consequences of single crossovers during tetrad formation in a reciprocal translocation heterozygote?
6. Give the gametic complement, in terms of acentrics, dicentrics, duplications, and deficiencies, when a three-strand double crossover occurs within a paracentric inversion loop.
7. In studying a new sample of fruit flies, a geneticist noted phenotypic variegation, semisterility, and the nonlinkage of previously linked genes. What probably caused this, and what cytological evidence would strengthen your hypothesis?
8. In a second sample of flies, the geneticist found a position effect and semisterility. The linkage groups were correct, but the order was changed and crossing over was suppressed. What probably caused this, and what cytological evidence would strengthen your hypothesis?

---

* Answers to selected exercises and problems are on page A-9.

**9.** Diagram the results of alternate segregation for a three-strand double crossover between a centromere and the cross center in a reciprocal translocation heterozygote.

**10.** A heterozygous plant *A B C D E/a b c d e* is testcrossed with an *a b c d e/a b c d e* plant. Only the following progeny appear.

*A B C D E/a b c d e*
*a b c d e/a b c d e*
*A b c d e/a b c d e*
*a B C D E/a b c d e*
*A B C D e/a b c d e*
*a b c d E/a b c d e*

What is unusual about the results? How can you explain them?

**11.** White eye color in *Drosophila* is an X-linked recessive trait. A wild-type male is irradiated and mated with a white-eyed female. Among the progeny is a white-eyed female.

**a.** Why is this result unexpected, and how could you explain it?

**b.** What type of progeny would you expect if this white-eyed female is crossed with a normal, nonirradiated male?

**12.** You are trying to locate an enzyme-producing gene in *Drosophila,* which you know is located on the third chromosome. You have five strains with deletions for different regions of the third chromosome (a slash — / — indicates a deleted region):

```
Normal    0   10   20   30   40   50   60    map units
Strain A  //////____________________________
Strain B  __//////////////////______________
Strain C  ____________////////////__________
Strain D  ___________________////////////___
Strain E  ________________________////////////
```

You cross each strain with wild-type flies and measure the amount of enzyme in the $F_1$ progeny. The results appear as follows. In what region is the gene located?

| Strain Crossed | Percentage of Wild-Type Enzyme Produced in $F_1$ Progeny |
|---|---|
| A | 100 |
| B | 45 |
| C | 54 |
| D | 98 |
| E | 101 |

**13.** Consider the following table, which shows the number of viable progeny produced by a plant under standard conditions. Provide an explanation for the results.

| $P_1$: | Strain A × Strain A | Strain B × Strain B | Strain A × Strain B |
|---|---|---|---|
| $F_1$: | 765 | 750 | 775 |
| $F_2$: | 712 | 783 | 416 |

**14.** The map position for three X-linked recessive genes in *Drosophila* (*v,* vermilion eyes; *m,* miniature wings; and *s,* sable body) is:

| *v* | *m* | *s* |
|---|---|---|
| 33.0 | 36.1 | 43.0 |

A wild-type male is X-rayed and mated to a vermilion, miniature, sable female. Among the progeny is a single vermilion-eyed, long-winged, tan-bodied female. The following shows the progeny when this female is mated with a *v m s* hemizygous male.

| Females | Males |
|---|---|
| 87 vermilion, miniature, sable | 89 vermilion, miniature, sable |
| 93 vermilion | 1 vermilion |

Explain these results by drawing a genetic map.

**15.** In *Drosophila,* recessive genes clot (*ct*) and black body (*b*) are located at 16.5 and 48.5 map units, respectively, on the second chromosome. In one cross, wild-type females that are $ct^+\ b^+/ct\ b$ are mated with *ct b/ct b* males. They produce these progeny:

| | |
|---|---|
| wild-type | 1,250 |
| clot, black | 1,200 |
| black | 30 |
| clot | 20 |

What is unusual about the results? How can you explain them?

**16.** You have four strains of *Drosophila* (1–4) that were isolated from different geographic regions. You compare the banding patterns of the second chromosome and obtain these results (each letter corresponds to a band):

(1) m n r q p o s t u v
(2) m n o p q r s t u v
(3) m n r q t s u p o v
(4) m n r q t s o p u v

If (3) is presumed to be the ancestral strain, in what order did the other strains arise?

**17.** In *Drosophila,* the recessive gene for white eyes is located near the tip of the X chromosome. A wild-type male is irradiated and mated with a white-eyed female. Among the progeny is one red-eyed male. How can you explain the red-eyed male, and how could you test your hypothesis?

**VARIATION IN CHROMOSOME NUMBER**

**18.** Is a tetraploid more likely to show irregularities in meiosis or mitosis? Explain. What about these processes in a triploid?

**19.** How many chromosomes would a human tetraploid have? How many chromosomes would a human monosomic have?

**20.** Do autopolyploids or allopolyploids experience more difficulties during meiosis? Do amphidiploids have more or less trouble than auto- or allopolyploids?

**21.** If a diploid species of $2n = 16$ hybridizes with one of $2n = 12$, and the resulting hybrid doubles its chromosome number to produce an allotetraploid (amphidiploid), how many chromosomes will it have? How many chromosomes will an allotetraploid have if both parent species had $2n = 20$?

**22.** If nondisjunction of the sex chromosomes occurs in a female at the second meiotic division, what type of eggs will arise?

**23.** How might an X0/XYY human mosaic arise? An XX/XXY mosaic? How might a trisomy 21 individual arise?

**24.** Plant species P has $2n = 18$, and species U has $2n = 14$. A fertile hybrid is found. How many chromosomes does it have?

**25.** A woman with normal vision whose father was color-blind mates with a man with normal vision. They have a color-blind daughter with Turner syndrome. In which parent did nondisjunction occur?

**26.** A color-blind man mates with a woman with normal vision whose father was color-blind. They have a color-blind son with Klinefelter syndrome. In which parent did nondisjunction occur?

**27.** Describe a genetic event that can produce an XYY man.

**28.** Chromosomal analysis of a spontaneously aborted fetus revealed that the fetus was 92,XXYY. Propose an explanation to account for this unusual karyotype.

## CRITICAL THINKING QUESTIONS

**1.** Various species in the grass genus *Bromus* have chromosome numbers of 14, 28, 42, 56, 70, 84, 98, and 112. What can you tell about the genetic relationships among these species and how they might have arisen?

**2.** There was a humorous television commercial in which someone accidentally discovered the desirability of combining chocolate and peanut butter. Could this combination be achieved by crossing peanut and cocoa plants?

*Suggested Readings for chapter 8 are on page B-4.*

# 9

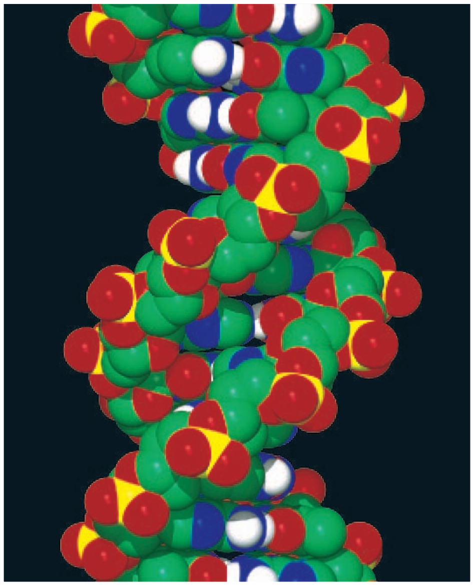
# CHEMISTRY OF THE GENE

## STUDY OBJECTIVES

1. To understand the properties that a genetic material must have 205
2. To examine the structure of DNA, the genetic material 211
3. To investigate the way in which DNA replicates 220

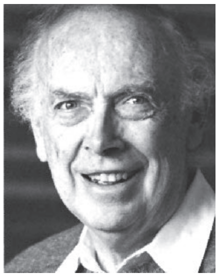
## STUDY OUTLINE

**In Search of the Genetic Material** 205
  Required Properties of a Genetic Material 205
  Evidence for DNA as the Genetic Material 209
**Chemistry of Nucleic Acids** 211
  Biologically Active Structure 214
  Requirements of Genetic Material 218
  Alternative Forms of DNA 219
**DNA Replication—The Process** 220
  The Meselson and Stahl Experiment 220
  Autoradiographic Demonstration of DNA Replication 222
**DNA Replication—The Enzymology** 225
  Continuous and Discontinuous DNA Replication 225
  The Origin of DNA Replication 229
  Events at the Y-Junction 232
  Supercoiling 234
  Termination of Replication 236
  DNA Partitioning in *E. coli* 238
**Replication Structures** 238
  Rolling-Circle Model 238
  D-Loop Model 238
**Eukaryotic DNA Replication** 238
*Summary* 240
*Solved Problems* 240
*Exercises and Problems* 241
*Critical Thinking Questions* 242
**Box 9.1 Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid** 206
**Box 9.2 Prions: The Biological Equivalent of Ice-Nine** 213
**Box 9.3 Multiple-Stranded DNA** 221

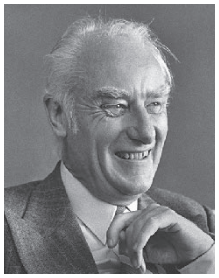


A computer-generated image of deoxyribonucleic acid, DNA. (© Professor K. Seddon & Dr. T. Evans/ Queen's University Belfast/SPL/Photo Researchers.)

In 1953, James Watson and Francis Crick published a two-page paper in the journal *Nature* entitled "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid." It began as follows: "We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest." This paper, which first put forth the correct model of DNA structure, is a milestone in the modern era of molecular genetics, compared by some to the work of Mendel and Darwin (box 9.1). (Watson, Crick, and X-ray crystallographer Maurice Wilkins won Nobel Prizes for this work; Rosalind Franklin, also an X-ray crystallographer, was acknowledged, posthumously, to have played a major role in the discovery of the structure of DNA.) Once the structure of the genetic material had been determined, an understanding of its method of replication and its functioning quickly followed.



James D. Watson (1928– ). (Cold Spring Harbor Laboratory Research Library Archives. Margot Bennet, photographer.)



Francis Crick (1916– ). (Reproduced by permission of Herb Weitman, Washington University, St. Louis, Missouri.)

Maurice H. F. Wilkins (1916– ). (Courtesy of Dr. Maurice H. F. Wilkins and Biophysics Department, King's College, London.)

## IN SEARCH OF THE GENETIC MATERIAL

This chapter begins a sequence of nine chapters on the molecular structure of the genetic material, its replication, its expression, and the control of its expression. In this chapter, we look at the evidence that DNA is the genetic material, the chemistry of DNA, and the way in which DNA replicates, including the general enzymatic processes. We look first at prokaryotic, then at eukaryotic, DNA replication. Note that we concentrate on the molecular structure of DNA because, generally, structure reveals function: molecules have shapes that define how they work.

### Required Properties of a Genetic Material

We begin with a look at the properties that a genetic material must have and review the evidence that nucleic acids make up the genetic material. To comprise the genes, DNA must carry the information to control the synthesis of the enzymes and proteins within a cell or organism; self-replicate with high fidelity, yet show a low level of mutation; and be located in the chromosomes.

#### Control of the Proteins

The growth, development, and functioning of a cell are controlled by the proteins within it, primarily its enzymes. Thus, the nature of a cell's phenotype is controlled by the protein synthesis within that cell. The genetic material must therefore determine the need for and effective amounts of the enzymes in a cell. For example, given inorganic salts and glucose, an *E. coli* cell can synthesize, through its enzyme-controlled biochemical pathways, all of the compounds it needs for growth, survival, and reproduction. In contrast, a mammalian red blood cell primarily produces hemoglobin.

At this point we need to review some basic information regarding enzymes. An enzyme is a protein that acts as a catalyst for a specific metabolic process without itself being markedly altered by the reaction. Most reactions that enzymes catalyze could occur anyway, but only under conditions too extreme to take place within living systems. For example, many oxidations occur naturally at high temperatures. Enzymes allow these reactions to occur within the cell by lowering the **free energy of activation ($\Delta G^{\ddagger}$)** of a particular reaction. In other words, an enzyme allows a reaction to take place without needing the boost in energy that heat usually supplies (fig. 9.1).

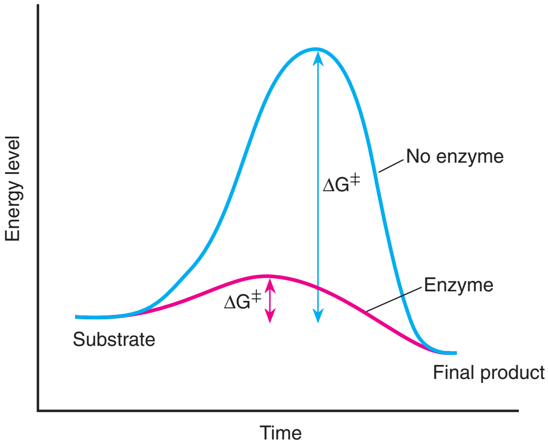


**Figure 9.1** An enzyme lowers the free energy of activation ($\Delta G^{\ddagger}$) for a particular reaction.

## BOX 9.1

## Historical Perspectives

### *Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid*

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey.[1] They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram [fig. 1]). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′, 5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration,' the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3.4 A in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

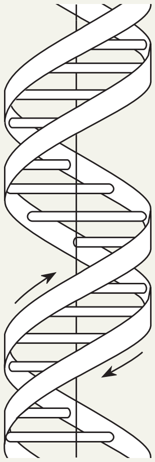


**Figure 1** This figure is purely diagrammatic. The two *ribbons* symbolize the two phosphate-sugar chains, and the *horizontal rods* represent the pairs of bases holding the chains together. The *vertical line* marks the fiber axis. (Reprinted with permission from *Nature,* Vol. 171, No. 4356. Watson and Crick, "Molecular Structure of Nucleic Acids," pp. 737–738. Copyright © 1953 Macmillan Magazines Limited.)

Most metabolic processes, such as the biosynthesis or degradation of molecules, occur in pathways, with enzyme facilitating each step in the pathway (see chapter 2). The metabolic pathway for the conversion of threonine into isoleucine (two amino acids) appears in figure 9.2. Each reaction product in the pathway is altered by an enzyme that converts it to the next product. The enzyme threonine dehydratase, for example, converts threonine into α-ketobutyric acid. Enzymes are composed of folded polymers of amino acids. The average protein is three hundred to five hundred amino acids long; only twenty naturally occurring amino acids are used in constructing these proteins. The sequence of amino acids determines the final structure of an enzyme. (We discuss the structure of proteins in more detail in chapter 11.) The genetic material determines the sequence of the amino acids.

The three-dimensional structure of enzymes permits them to perform their function. An enzyme combines with its substrate or substrates (the molecules it works on) at a part of the enzyme called the **active site** (fig. 9.3). The substrates "fit" into the active site, which has a shape that allows only the specific substrates to enter. This view of the way an enzyme interacts with its substrates is called the *lock-and-key model* of enzyme functioning.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations), it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid. It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereo-chemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their coworkers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

*J. D. Watson*
*F. H. C. Crick*

*Medical Research Council Unit for the Study of Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge. April 2.*

Reprinted by permission from *Nature,* 171:737–38. Copyright © 1953 Macmillan Magazines Ltd.

1. Pauling, L., and Corey, R. B., *Nature,* 171, 346 (1953); *Proc. U.S. Nat. Acad. Sci.,* 39, 84 (1953).
2. Furberg, S., *Acta Chem. Scand.,* 6, 634 (1952).
3. Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta,* 9, 402 (1952).
4. Wyatt, G. R., *J. Gen. Physiol.,* 36, 201 (1952).
5. Astbury, W. T., *Symp. Soc. Exp. Biol. 1, Nucleic Acid* 66 (Camb. Univ. Press, 1947).
6. Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys.* Acta, 10, 192 (1953).

When the substrates are in their proper position in the active site of the enzyme, the particular reaction that the enzyme catalyzes takes place. The reaction products then separate from the enzyme and leave it free to repeat the process. Enzymes can work at phenomenal speeds. Some can catalyze as many as a million reactions per minute.

Not all of the cell's proteins function as catalysts. Some are structural proteins, such as keratin, the main component of hair. Other proteins are regulatory—they control the rate at which other enzymes work. Still others are involved in different functions; albumins, for example, help regulate the osmotic pressure of blood.

### *Replication*

The genetic material must be capable of precisely directing its own replication so that every daughter cell receives an exact copy. Some **mutability,** or the ability to change, is also required, because we know that the genetic material has changed, or evolved, over the history of life on earth. In their 1953 paper, Watson and Crick had already worked out the replication process based on the structure of DNA. The fidelity of the replication process is so great that the error rate is only about one in a billion.

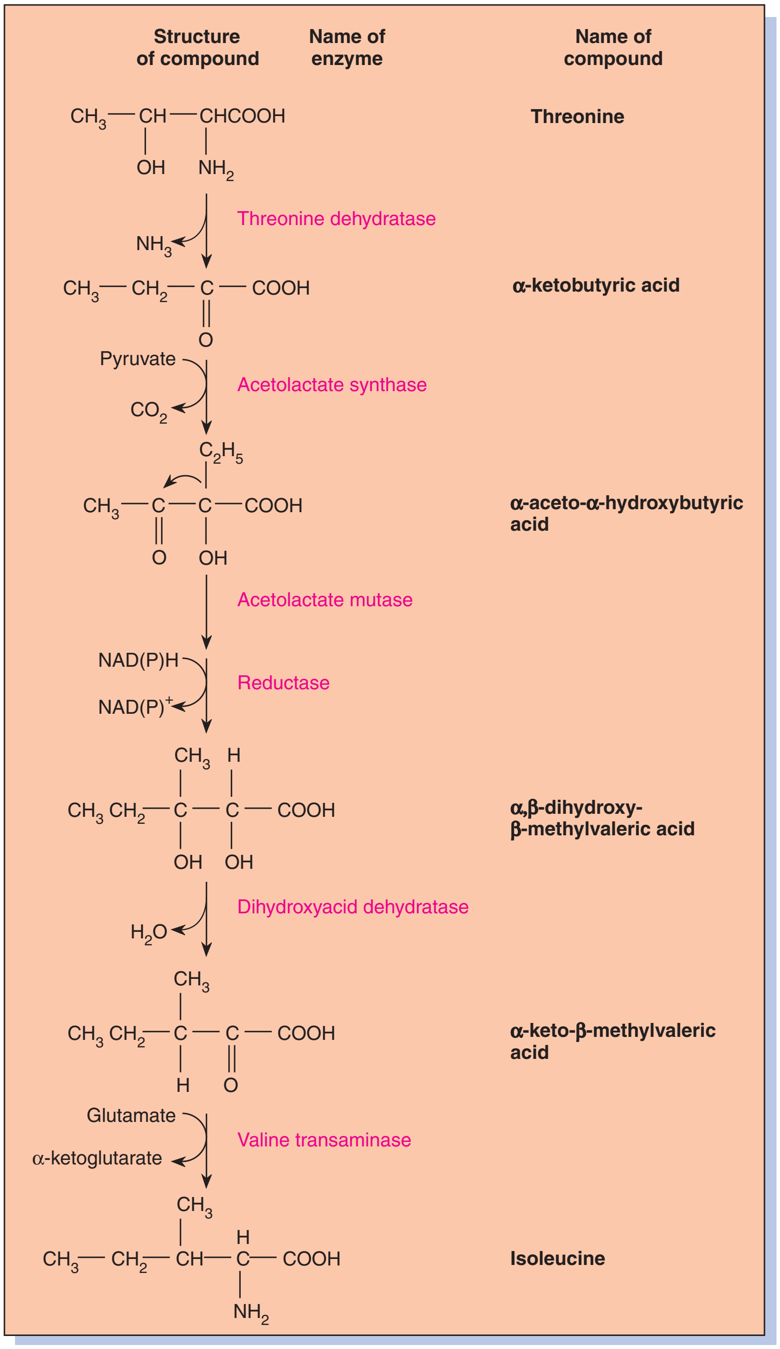

Structure of compound
Name of enzyme
Name of compound
Threonine
Threonine dehydratase
$NH_3$
α-ketobutyric acid
Pyruvate
Acetolactate synthase
$CO_2$
α-aceto-α-hydroxybutyric acid
Acetolactate mutase
NAD(P)H
Reductase
$NAD(P)^+$
α,β-dihydroxy-β-methylvaleric acid
Dihydroxyacid dehydratase
$H_2O$
α-keto-β-methylvaleric acid
Glutamate
Valine transaminase
α-ketoglutarate
Isoleucine

**Figure 9.2** Metabolic pathway of conversion of the amino acid threonine into isoleucine.

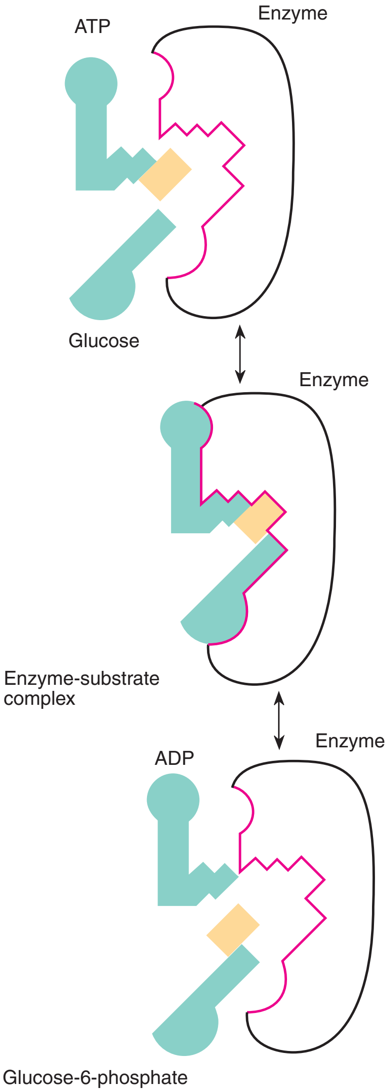
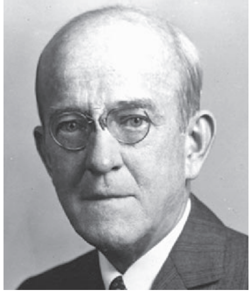

ATP
Enzyme
Glucose
Enzyme
Enzyme-substrate complex
ADP
Enzyme
Glucose-6-phosphate

**Figure 9.3** The active site of an enzyme recognizes a specific substance. In this case, ATP plus glucose is converted into ADP and glucose-6-phosphate by the enzyme hexokinase. The active site is diagrammed in *red*. The terminal phosphate group of ATP is *tan*.

### *Location*

It has been known since the turn of the century that genes, the discrete functional units of genetic material, are located in chromosomes within the nuclei of eukaryotic cells: the way chromosomes behave during the cellular division stages of mitosis and meiosis mimics the behavior of genes. Thus, the genetic material in eukaryotes must be a part of the chromosomes.

For a long time, proteins were considered the most probable genetic material because they have the neces-



Oswald T. Avery (1877–1955). (Courtesy of the National Academy of Sciences.)

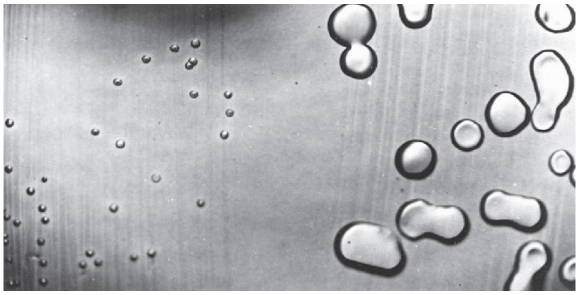


**Figure 9.4** Petri plate with smooth and rough colonies of *Streptococcus pneumoniae.* R (rough) strain colonies appear on the *left* and S (smooth) colonies on the *right* on the same agar. Magnification 3.5×. (O. T. Avery, C. M. Macleod, and M. McCarty, "Studies on the chemical nature of the substance inducing transformation of pneumococcal types." Reproduced from the *Journal of Experimental Medicine* 79 (1944):137–58, fig. 1 by copyright permission of the Rockefeller University Press. Reproduced by permission. Photograph made by Mr. Joseph B. Haulenbeek.)

sary molecular complexity. The twenty naturally occurring amino acids can be combined in an almost unlimited variety, creating thousands and thousands of different proteins. The first proof that the genetic material is deoxyribonucleic acid (DNA) came in 1944 from Oswald Avery and his colleagues. The Watson and Crick model in 1953 ended a period when many thought DNA was the genetic material, but its structure was unknown.

## Evidence for DNA as the Genetic Material

### Transformation

In 1928, F. Griffith reported that heat-killed bacteria of one type could "transform" living bacteria of a different type. Griffith demonstrated this transformation using two strains of the bacterium *Streptococcus pneumoniae.* One strain (S) produced smooth colonies on media in a petri plate because the cells had polysaccharide capsules. It caused a fatal bacteremia (bacterial infection) in mice. Another strain (R), which lacked polysaccharide capsules, produced rough colonies on petri plates (fig. 9.4); it did not have a pathological effect on mice. Bacteria of the rough strain are engulfed by the mice's white blood cells; bacteria of the virulent smooth strain survive because their polysaccharide coating protects them.

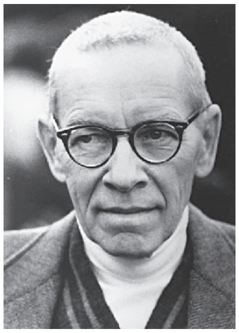
Griffith found that neither heat-killed S-type nor live R-type cells, by themselves, caused bacteremia in mice. However, if he injected a mixture of live R-type and heat-killed S-type cells into mice, the mice developed a bacteremia identical to that caused by living S-type cells (fig. 9.5). Thus, something in the heat-killed S cells transformed the R-type bacteria into S-type cells.

In 1944, Oswald Avery and two of his associates, C. MacLeod and M. McCarty, reported the nature of the transforming substance. Avery and his colleagues did their work **in vitro** (literally, in glass), using colony morphology on culture media rather than bacteremia in mice as evidence of transformation. They ruled out proteins, carbohydrates, and lipids by their extraction procedure, by the chemical analysis of the transforming material, and by demonstrating that the only enzymes that destroyed the transforming ability were enzymes that destroyed DNA. This study provided the first experimental evidence that DNA was the genetic material: DNA transformed R-type bacteria into S-type bacteria.

### Phage Labeling

Valuable information about the nature of the genetic material has also come from viruses. Of particular value are studies of bacterial viruses—the bacteriophages, or phages. Since phages consist only of nucleic acid surrounded by protein, they lend themselves nicely to the determination of whether the protein or the nucleic acid is the genetic material.

A. D. Hershey and M. Chase published, in 1952, the results of research that supported the notion that DNA is the genetic material and, in the process, helped to



A. D. Hershey (1908–1997). (Courtesy of Dr. A. D. Hershey.)

Martha Chase. (Courtesy of Cold Springs Harbor Laboratory Archives.)

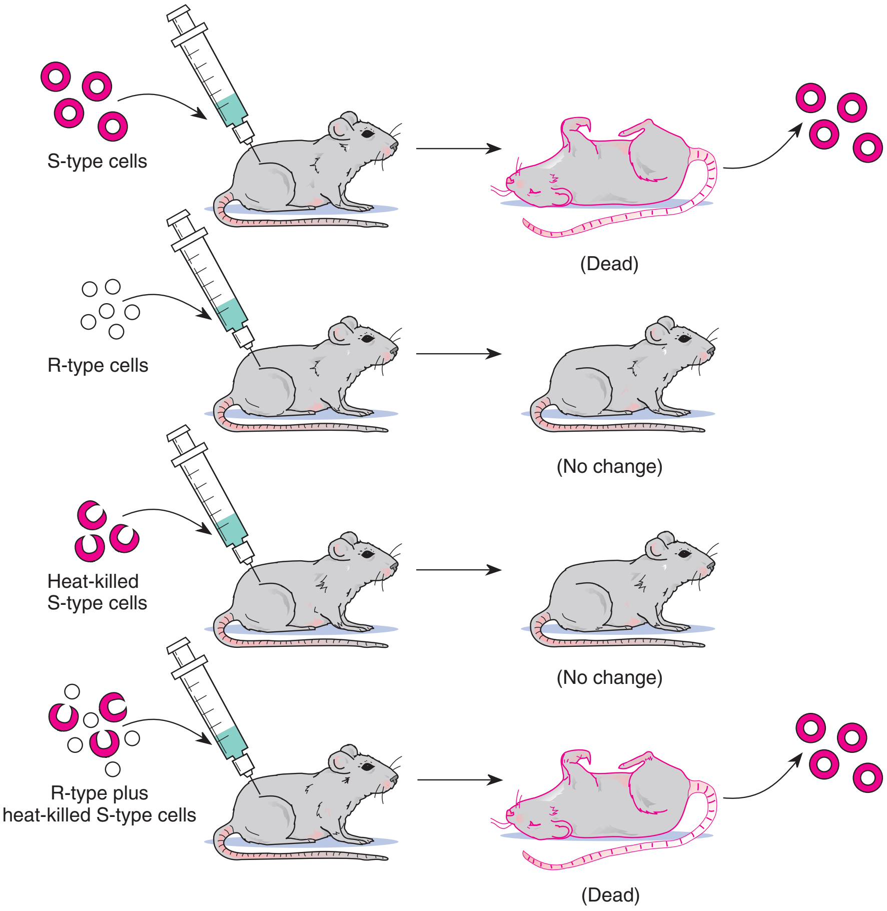


**Figure 9.5** Griffith's experiment with *Streptococcus.* S-type cells will kill mice; so will heat-killed S-type cells injected with live R-type cells. S-type cells are recovered from dead mice in both cases.

explain the nature of the viral infection process. Since all nucleic acids contain phosphorus, whereas proteins do not, and since most proteins contain sulfur (in the amino acids cysteine and methionine), whereas nucleic acids do not, Hershey and Chase designed an experiment using radioactive isotopes of sulfur and phosphorus to keep separate track of the viral proteins and nucleic acids during the infection process. They used the T2 bacteriophage and the bacterium *Escherichia coli.* The phages were labeled by having them infect bacteria growing in culture medium containing the radioactive isotopes $^{35}$S or $^{32}$P. Hershey and Chase then proceeded to identify the material injected into the cell by phages attached to the bacterial wall.

When $^{32}$P-labeled phages were mixed with unlabeled *E. coli* cells, Hershey and Chase found that the $^{32}$P label entered the bacterial cells and that the next generation of phages that burst from the infected cells carried a significant amount of the $^{32}$P label. When $^{35}$S-labeled phages were mixed with unlabeled *E. coli,* the researchers found that the $^{35}$S label stayed outside the bacteria for the most part. Hershey and Chase thus demonstrated that the outer protein coat of a phage does not enter the bacterium it infects, whereas the phage's inner material, consisting of DNA, does enter the bacterial cell (fig. 9.6). Since the DNA is responsible for the production of the new phages during the infection process, the DNA, not the protein, must be the genetic material.

### RNA as Genetic Material

In some viruses, RNA (ribonucleic acid) is the genetic material. The tobacco mosaic virus that infects tobacco plants consists only of RNA and protein. The single, long RNA molecule is packaged within a rodlike structure formed by over two thousand copies of a single protein. No DNA is present in tobacco mosaic virus particles (fig. 9.7*a*). In 1955, H. Fraenkel-Conrat and R. Williams

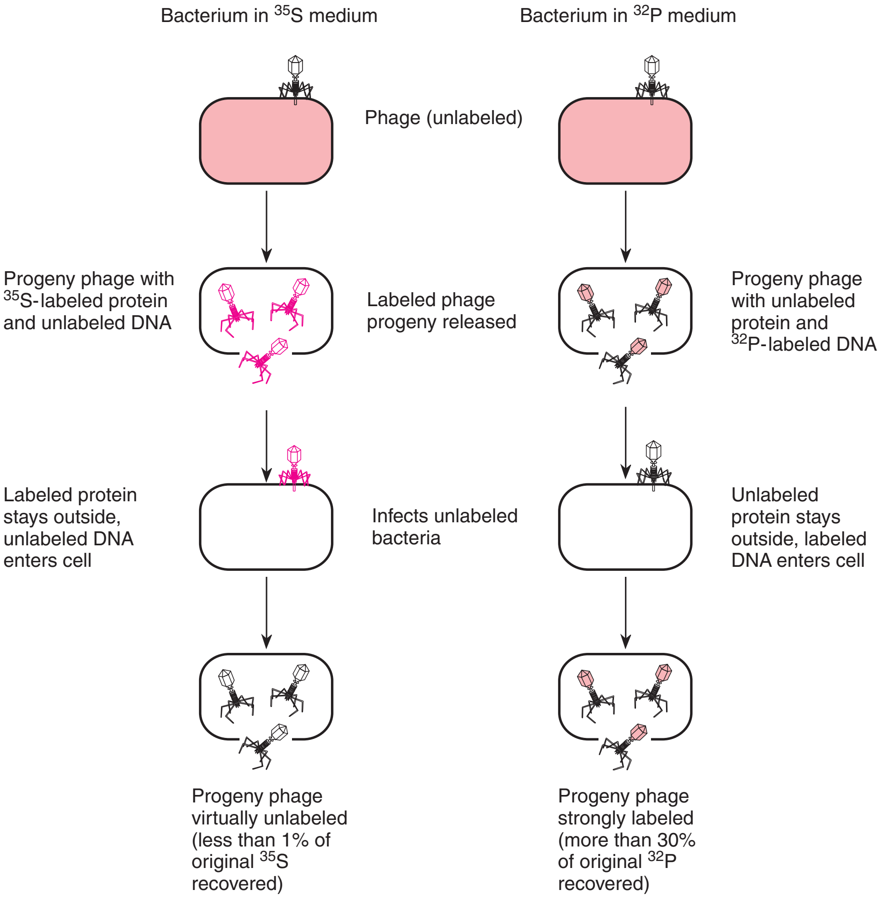

Bacterium in $^{35}S$ medium
Bacterium in $^{32}P$ medium
Phage (unlabeled)
Progeny phage with $^{35}S$-labeled protein and unlabeled DNA
Labeled phage progeny released
Progeny phage with unlabeled protein and $^{32}P$-labeled DNA
Labeled protein stays outside, unlabeled DNA enters cell
Infects unlabeled bacteria
Unlabeled protein stays outside, labeled DNA enters cell
Progeny phage virtually unlabeled (less than 1% of original $^{35}S$ recovered)
Progeny phage strongly labeled (more than 30% of original $^{32}P$ recovered)

**Figure 9.6** The Hershey and Chase experiments using $^{35}S$-labeled and $^{32}P$-labeled T2 bacteriophages. The nucleic acid label ($^{32}P$) enters the *E. coli* bacteria during infection; the protein label ($^{35}S$) does not.

showed that a virus can be separated, in vitro, into its component parts and reconstituted as a viable virus. This finding led Fraenkel-Conrat and B. Singer to reconstitute tobacco mosaic virus with parts from different strains (fig. 9.7*b*). For example, they combined the RNA from the common tobacco mosaic virus with the protein from the masked (M) strain of tobacco mosaic virus. They then made the reciprocal combination of common-type protein and M-type RNA. In both cases, the tobacco mosaic virus produced during the process of infection was the type associated with the RNA, not with the protein. Thus, it was the nucleic acid (RNA in this case) that was the genetic material. Subsequently, scientists rubbed pure tobacco mosaic virus RNA into plant leaves. Normal infection and a new generation of typical, protein-coated tobacco mosaic virus resulted, confirming RNA as the genetic material for this virus.

We thus conclude that DNA is the genetic material. In the few viruses that do not have DNA, RNA serves as the genetic material. The only exception to these statements is one type of disease that is transmitted by a protein without accompanying DNA or RNA (box 9.2).

## CHEMISTRY OF NUCLEIC ACIDS

Having identified the genetic material as the nucleic acid DNA (or RNA), we proceed to examine the chemical structure of these molecules. Their structure will tell us a good deal about how they function.

Nucleic acids are made by joining **nucleotides** in a repetitive way into long, chainlike polymers. Nucleotides are made of three components: phosphate, sugar, and a nitrogenous base (table 9.1 and fig. 9.8). When incorporated into a nucleic acid, a nucleotide contains one of each of the three components. But, when free in the cell

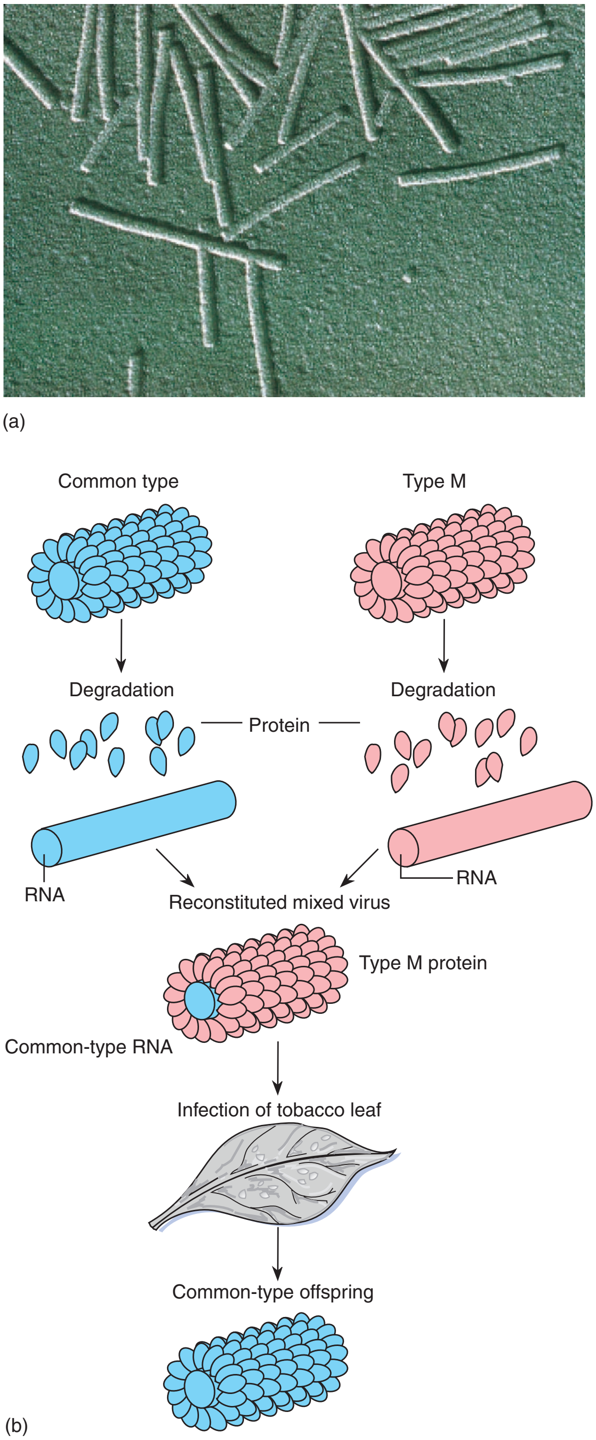

(a)
Common type
Type M
Degradation
Degradation
Protein
RNA
RNA
Reconstituted mixed virus
Type M protein
Common-type RNA
Infection of tobacco leaf
Common-type offspring
(b)

**Figure 9.7** (*a*) Electron micrograph of tobacco mosaic virus. Magnification 37,428×. (*b*) Reconstitution experiment of Fraenkel-Conrat and Singer. The nucleic acid (RNA), not the protein component of the virus, controls inheritance.

([*a*] © Biology Media/Photo Researchers, Inc.)

**Table 9.1 Components of Nucleic Acids**

| | | | Base | |
|---|---|---|---|---|
| | **Phosphate** | **Sugar** | **Purines** | **Pyrimidines** |
| DNA | Present | Deoxyribose | Guanine<br>Adenine | Cytosine<br>Thymine |
| RNA | Present | Ribose | Guanine<br>Adenine | Cytosine<br>Uracil |

pool, nucleotides usually occur as triphosphates. The energy held in the extra phosphates is used, among other purposes, to synthesize the polymer. A **nucleoside** is a sugar-base compound. Nucleotides are therefore nucleoside phosphates (fig. 9.9). (Note that ATP, adenosine triphosphate, the energy currency of the cell, is a nucleoside triphosphate.)

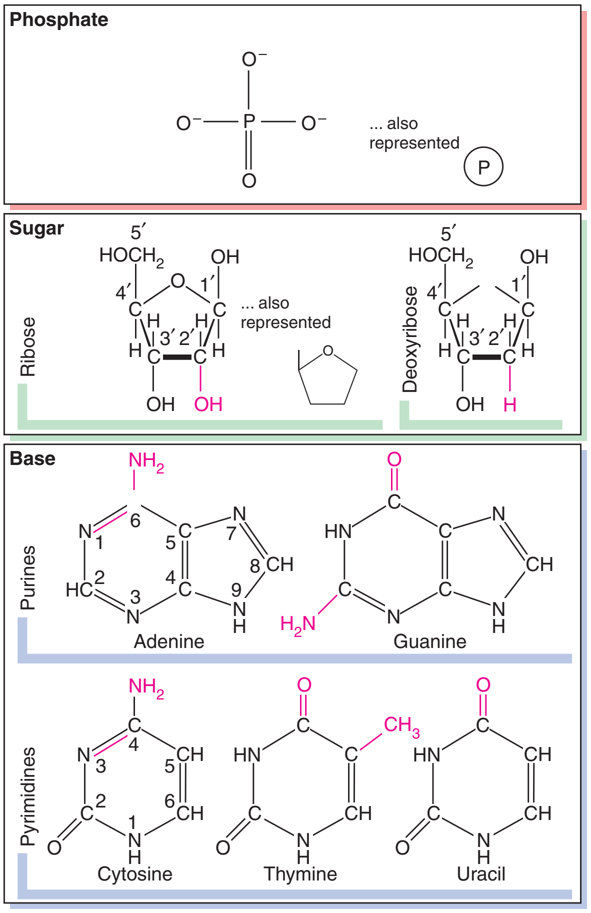

Phosphate
... also represented
P
Sugar
Ribose
... also represented
Deoxyribose
Base
Purines
Adenine
Guanine
Pyrimidines
Cytosine
Thymine
Uracil

**Figure 9.8** Components of nucleic acids: phosphate, sugars, and bases. Primes are used to number the ring positions in the sugars to differentiate them from the ring positions in the bases.

## BOX 9.2

### Biomedical Applications

### *Prions: The Biological Equivalent of Ice-Nine*

Without exception, the genetic material is either DNA or RNA; it is RNA only in a few viruses. Since virtually all transmissible diseases are of bacterial or viral origin, this means that transmissible diseases are also caused by organisms with DNA or RNA as their genetic material. However, in one interesting situation, a transmissible disease appears to be caused by an agent without genetic material. Four human neurological diseases and six similar animal diseases are caused, we believe, by proteins without DNA or RNA. (Two conditions in yeast are probably caused in a similar way.) The human diseases are kuru, Creutzfeldt-Jakob disease, Gerstmann-Sträussler-Scheinker syndrome, and a recently discovered fatal familial insomnia. The animal diseases are scrapie (sheep and goats), four encephalopathies (bovine, feline, ungulate, and mink), and chronic wasting disease (deer and elk). All of these diseases are extremely slow to develop, all are fatal, and all are believed to be caused either by the ingestion of a protein from an infected individual or from a mutation of the normal gene. None of the diseases as yet has a cure, and the mechanism of action is not completely understood.

The diseases appear to be caused by a protein, similar to one normally produced in the brain of healthy individuals. The term **prion** (taken from *pr*oteinaceous *i*nfectious particle) has been given to these agents by Stanley Prusiner at the University of California in San Francisco, a 1997 Nobel laureate. He, along with colleagues, isolated the prion protein (PrP) and recently located the gene that codes for the protein on the short arm of chromosome 20. In addition to the infective form, a familial (inherited) form of these diseases can result from a mutation of the gene that codes for the prion protein active in normal individuals (probably at least all mammals). The normal protein is termed $\mathrm{PrP}^{\mathrm{C}}$, and the mutated form is referred to as $\mathrm{PrP}^{\mathrm{Sc}}$. Normally, $\mathrm{PrP}^{\mathrm{C}}$ is a glycoprotein found on the membrane surface of the cells of the brain and some other tissues.

Although no cures exist for these diseases, kuru, at least, seems to be almost eradicated. It was found only among people in part of New Guinea who practiced cannibalism. Once the people stopped this practice, the spread of the disease also ceased; kuru does not seem to be generated to any major extent by mutation. By controlling feeding practices, it is believed, bovine spongiform encephalopathy will also disappear. In the past, cows were fed protein supplements contaminated by material from infected animals.

In England, a recent epidemic of bovine spongiform encephalopathy (BSE, or mad cow disease) peaked in 1992–1993, affecting over 160,000 cattle. At least fourteen cases of a variant of Creutzfeldt-Jakob disease in people in England and France were attributed to eating affected beef, creating a panic in England. With a change away from using animal matter in cattle feed and a culling of cattle herds, the epidemic has ended. However, new human cases may show up in the future owing to the long incubation period of this prion disease.

The obvious question is, how does a protein that does not appear to contain genetic material cause a transmissible disease when ingested? Prusiner has suggested several mechanisms that would allow an infective protein to induce copies of the normal protein to become infective. One of these mechanisms involves a cascade in which an infective $\mathrm{PrP}^{\mathrm{Sc}}$ binds with a normal $\mathrm{PrP}^{\mathrm{C}}$, resulting in two infective $\mathrm{PrP}^{\mathrm{Sc}}$ proteins. From this, one produces two, two produce four, four produce eight, and so on. As Nancy Touchette, writing in *The Journal of NIH Research,* pointed out, this is the way Kurt Vonnegut described the behavior of the mythical ice-nine in his 1963 book, *Cat's Cradle.* In this fictional account, a single seed caused all of the water on earth, by a chain reaction cascade, to form into a novel type of ice. We have not yet resorted to science fiction to answer the mystery of prion function; however, it seems reasonable to guess that an eventual understanding of the mechanism of prion function will provide us with a biological novelty.

---

The sugars differ only in the presence (ribose in RNA) or absence (deoxyribose in DNA) of an oxygen in the $2'$ position. (The carbons of the sugars are numbered $1'$ to $5'$. The primes are used to avoid confusion with the numbering system of the bases; see fig. 9.8.) DNA and RNA both have four bases (two **purines** and two **pyrimidines**) in their nucleotide chains. Both molecules have the purines **adenine** and **guanine** and the pyrimidine **cytosine.** DNA has the pyrimidine **thymine;** RNA has the pyrimidine **uracil.** Thus, three of the nitrogenous bases are found in both DNA and RNA, whereas thymine is unique to DNA, and uracil is unique to RNA.

A nucleotide is formed in the cell when a base attaches to the $1'$ carbon of the sugar and a phosphate attaches to

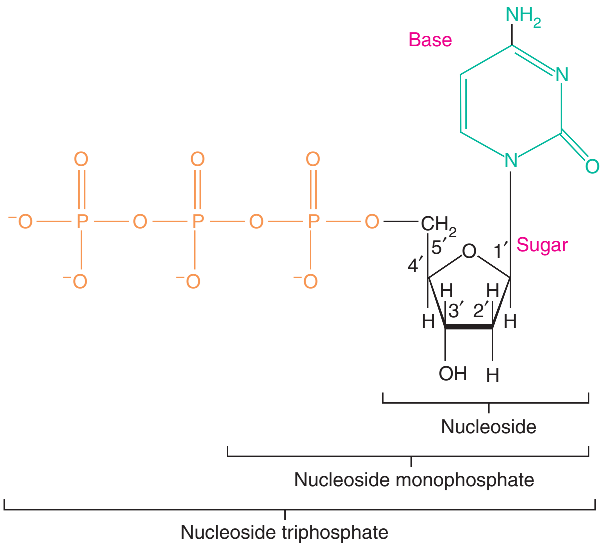

Base
$NH_2$
N
O
Sugar
$^-O$
P
O
$CH_2$
5′
4′
1′
3′
2′
H
OH
Nucleoside
Nucleoside monophosphate
Nucleoside triphosphate

**Figure 9.9** The structure of a nucleoside and two nucleotides: a nucleoside monophosphate and a nucleoside triphosphate.

the 5′ carbon of the same sugar (fig. 9.10); the nucleotide takes its name from the base (table 9.2). Nucleotides are linked together **(polymerized)** by the formation of a bond between the phosphate at the 5′ carbon of one nucleotide and the hydroxyl (OH) group at the 3′ carbon of an adjacent molecule. Very long strings of nucleotides can be polymerized by this **phosphodiester bonding** (fig. 9.11).

## Biologically Active Structure

Although the identities of the nucleotides that polymerized to form a strand of DNA or RNA were known, the actual structures of these nucleic acids when they function as the genetic material remained unknown until 1953. The general feeling was that the biologically active structure of DNA was more complex than a single string of nucleotides linked together by phosphodiester bonds, and that several interacting strands were involved. In 1953, Linus Pauling, a Nobel laureate who had discovered the

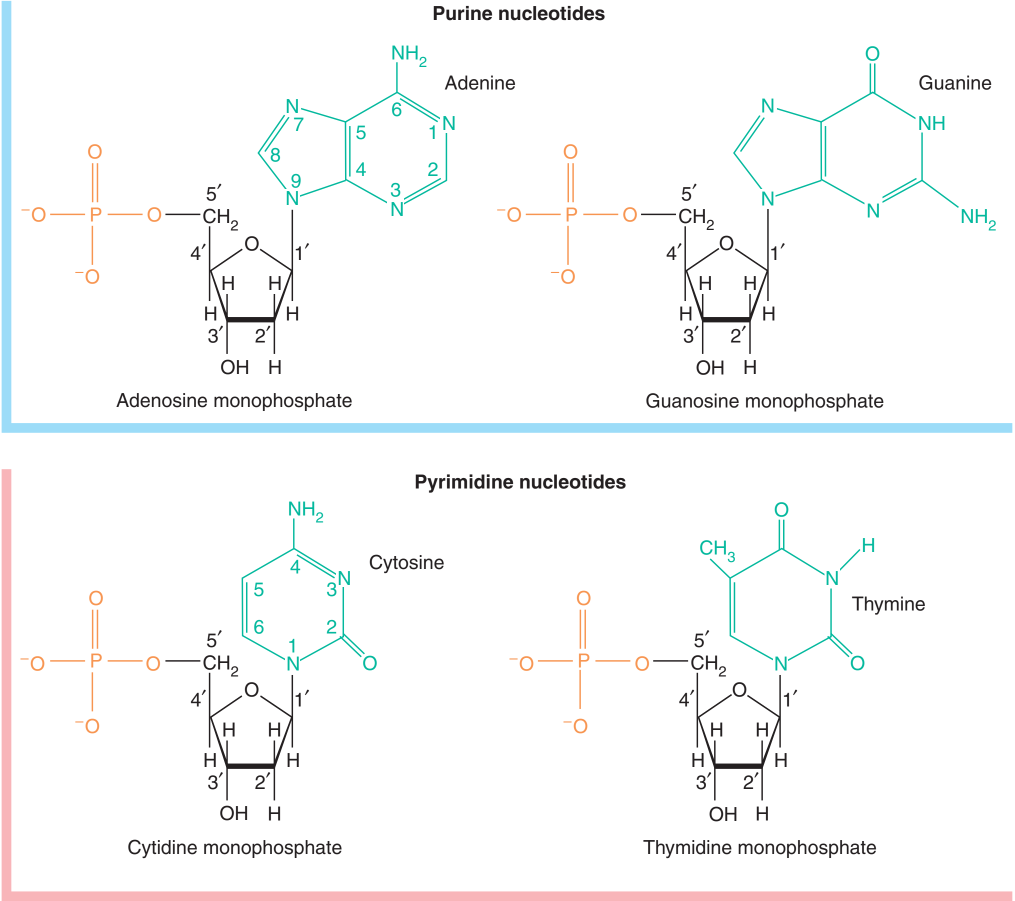

Purine nucleotides
Adenine
Guanine
$NH_2$
N
NH
O
$^-O$
P
$CH_2$
5′
4′
1′
3′
2′
H
OH
1
2
3
4
5
6
7
8
9
Adenosine monophosphate
Guanosine monophosphate
Pyrimidine nucleotides
Cytosine
Thymine
$CH_3$
Cytidine monophosphate
Thymidine monophosphate

**Figure 9.10** Structure of the four deoxyribose nucleotides.

**Table 9.2** Nucleotide Nomenclature

| | | Abbreviation | | | | | |
|---|---|---|---|---|---|---|---|
| | | Monophosphate | | Diphosphate | | Triphosphate | |
| **Base** | **Nucleotide (nucleoside monophosphate)** | **Ribose** | **Deoxyribose** | **Ribose** | **Deoxyribose** | **Ribose** | **Deoxyribose** |
| Guanine | Guanosine monophosphate | GMP | | GDP | | GTP | |
| | Deoxyguanosine monophosphate | | dGMP | | dGDP | | dGTP |
| Adenine | Adenosine monophosphate | AMP | | ADP | | ATP | |
| | Deoxyadenosine monophosphate | | dAMP | | dADP | | dATP |
| Cytosine | Cytidine monophosphate | CMP | | CDP | | CTP | |
| | Deoxycytidine monophosphate | | dCMP | | dCDP | | dCTP |
| Thymine | Deoxythymidine monophosphate | | dTMP | | dTDP | | dTTP |
| Uracil | Uridine monophosphate | UMP | | UDP | | UTP | |

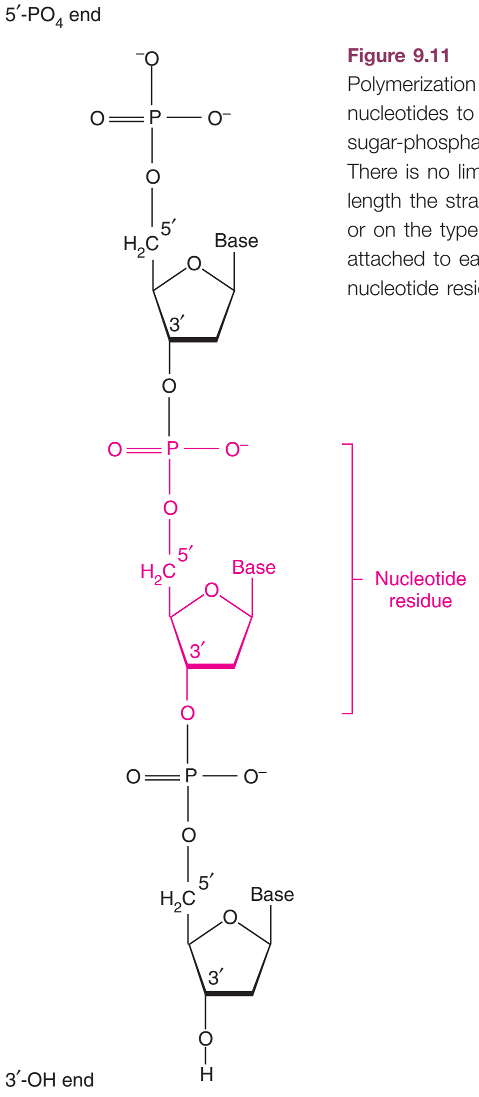


**Figure 9.11** Polymerization of adjacent nucleotides to form a sugar-phosphate strand. There is no limit to the length the strand can be or on the type of base attached to each nucleotide residue.

helical structure of proteins, was investigating a three-stranded structure for the genetic material, whereas Watson and Crick had decided that a two-stranded structure was more consistent with available evidence. Three lines of evidence directed Watson and Crick: the chemical nature of the components of DNA, X-ray crystallography, and Chargaff's ratios.

### *DNA X-Ray Crystallography*

All the time Watson and Crick were studying DNA structure, Maurice Wilkins, Rosalind Franklin, and their colleagues were using **X-ray crystallography** to analyze the structure of DNA. The molecules in a crystal are arranged in an orderly way, so that when a beam of X rays is aimed at the crystal, the beam scatteres in an orderly fashion. The scatter pattern can be recorded on photographic film or computer-controlled devices. The nature of this pattern depends on the structure of the crystal. The cross in the center of the photograph in figure 9.12 indicates that the molecule is a helix; the dark areas at the top and bottom come from the bases, stacked perpendicularly to the main axis of

Rosalind E. Franklin (1920–1958). (Courtesy of Cold Spring Harbor Laboratory.)

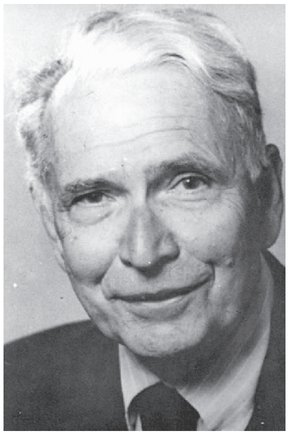
the molecule. This image of the DNA molecule stimulated Watson and Crick's understanding of its structure.

### *Chargaff's Ratios*

Until Erwin Chargaff's work, scientists had labored under the erroneous **tetranucleotide hypothesis.** This hypothesis proposed that DNA was made up of equal quantities of the four bases; therefore, a subunit of this DNA consisted of one copy of each base. Chargaff carefully analyzed the base composition of DNA in various species (table 9.3). He found that although the relative amount of a given nucleotide differs among species, the amount of adenine equaled that of thymine and the amount of guanine equaled that of cytosine. That is, in the DNA of all the organisms studied, a 1:1 correspondence exists between the purine and pyrimidine bases. This is known as **Chargaff's rule.** Chargaff's observations disproved the tetranucleotide hypothesis; the four bases of DNA did not occur in a 1:1:1:1 ratio. His results gave insight to Watson and Crick in the development of their model.

### *The Watson-Crick Model*

With the information available, Watson and Crick began constructing molecular models. They found that a possible structure for DNA was one in which two helices coiled around one another (a **double helix**), with the sugar-phosphate backbones on the outside and the bases on the inside. This structure would fit the dimensions X-ray crystallography had established for DNA if the bases from the two strands were opposite each other and formed "rungs" in a helical "ladder" (fig. 9.13). The diameter of the helix could only be kept constant at about 20 Å (10 angstrom units = 1 nanometer) if one purine and one pyrimidine base made up each rung. Two purines per rung would be too big, and two pyrimidines would be too small.



Erwin Chargaff (1905– ). (Courtesy of Dr. Erwin Chargaff.)

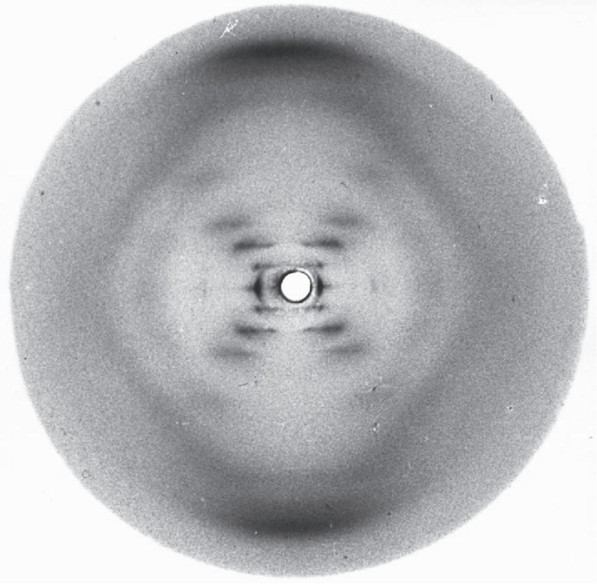


**Figure 9.12** Scatter pattern of a beam of X rays passed through crystalline DNA. (Source: Reprinted by permission from R. E. Franklin and R. Gosling, "Molecular configuration in sodium thymonucleate," *Nature* 171:740–41. Copyright 1953 by Macmillan Journals Limited.)

After further experimentation with models, Watson and Crick found that the hydrogen bonding necessary to form the rungs of their helical ladder could occur readily between certain base pairs, the pairs that Chargaff found in equal frequencies. (Hydrogen bonds are very weak bonds in which two electronegative atoms, such as O and N, share a hydrogen atom between them. They have 3 to 5% of the strength of a covalent bond.) Thermodynamically stable

**Table 9.3** **Percentage Base Composition of Some DNAs**

| Species | Adenine | Thymine | Guanine | Cytosine |
|---|---|---|---|---|
| Human being (liver) | 30.3 | 30.3 | 19.5 | 19.9 |
| *Mycobacterium tuberculosis* | 15.1 | 14.6 | 34.9 | 35.4 |
| Sea urchin | 32.8 | 32.1 | 17.7 | 18.4 |

*Source:* From E. Chargaff and J. Davidson, *The Nucleic Acids,* Academic Press, 1955.

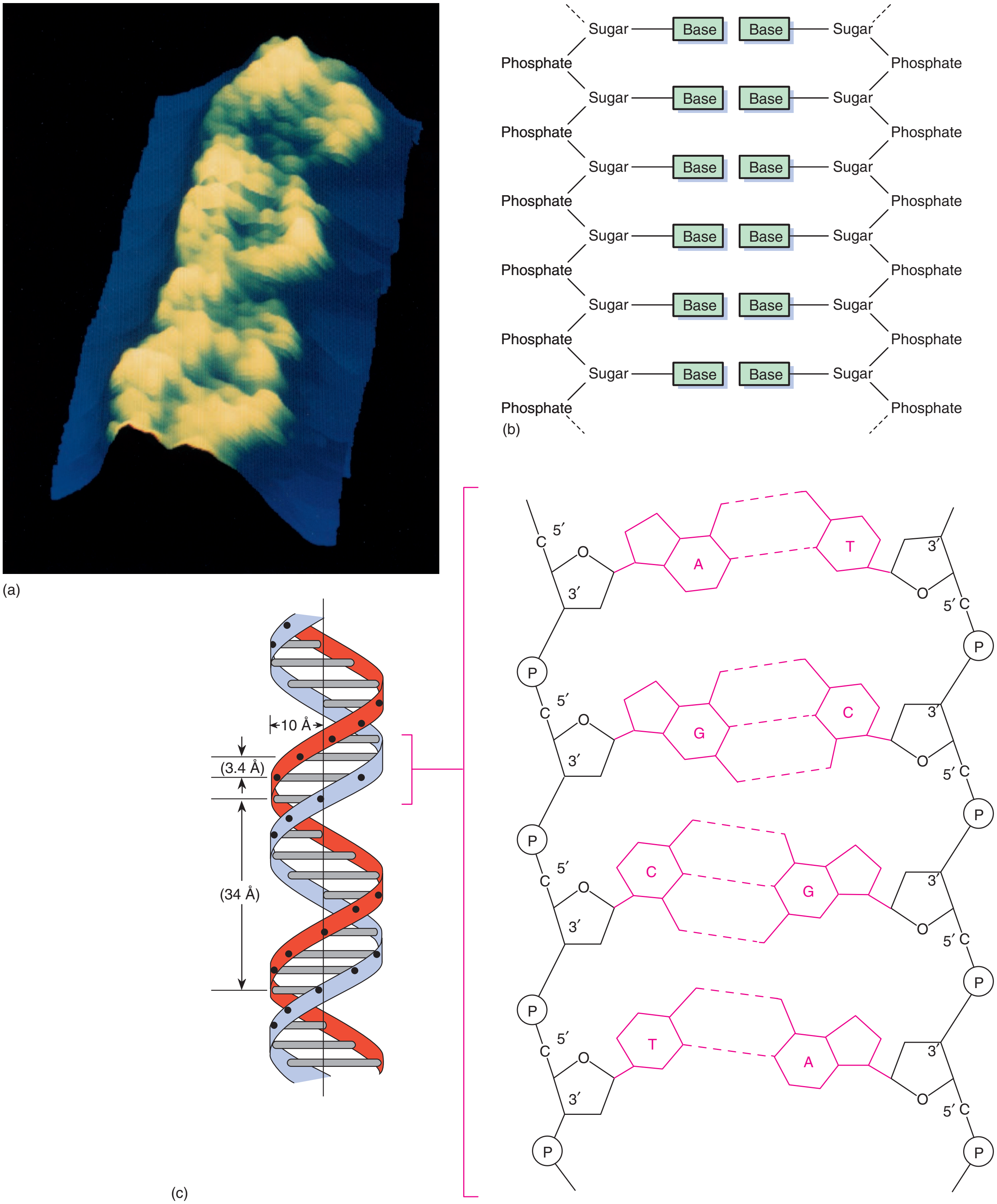

(a)
Sugar
Base
Base
Sugar
Phosphate
Phosphate
(b)
10 Å
(3.4 Å)
(34 Å)
(c)
5′
C
O
3′
P
A
T
G
C
C
G
T
A

**Figure 9.13** Double helical structure of DNA. (*a*) DNA magnified twenty-five million times by scanning tunneling microscopy. (*b*) Component parts. (*c*) Line drawing. ([*a*] © John D. Baldeschweiler.)

hydrogen bonding occurs between thymine and adenine and between cytosine and guanine (fig. 9.14). The relationship is one of **complementarity.** There are two hydrogen bonds between adenine and thymine and three between cytosine and guanine.

Another point about DNA structure relates to the **polarity** that exists in each strand. That is, one end of a DNA strand has a 5′ phosphate and the other end has a 3′ hydroxyl group. Watson and Crick found that hydrogen bonding would occur if the polarity of the two strands ran in opposite directions; that is, if the two strands were **antiparallel** (fig. 9.15).

### *DNA Denaturation*

Denaturation studies indicated that the hydrogen bonding in DNA occurs in the way Watson and Crick suggested. Hydrogen bonds, although individually very weak, give structural stability to a molecule in large enough numbers. However, the hydrogen bonds can be broken and the DNA strands separated when the DNA molecule is heated in water. At a certain point, the thermal agitation overcomes the hydrogen bonding, and the molecule becomes **denatured** (or "melts"). It is logical that the more hydrogen bonds DNA contains the higher the temperature needed to denature it. It thus follows that since a G-C (guanine-cytosine) base pair has three hydrogen bonds to every two in an A-T (adenine-thymine) base pair, the higher the G-C content in a given molecule of DNA, the higher the temperature required to denature it. This relationship exists (fig. 9.16).

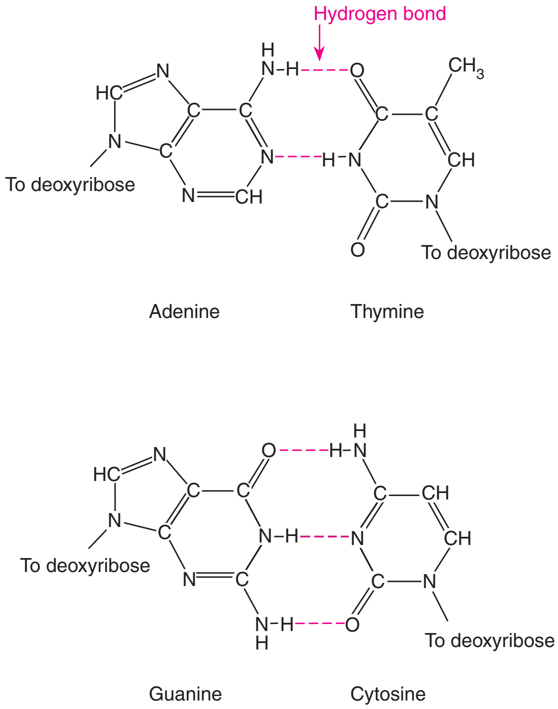


**Figure 9.14** Hydrogen bonding between the nitrogenous bases in DNA.

## *Requirements of Genetic Material*

Let us now return briefly to the requirements we have said a genetic material needs to meet: (1) control of protein synthesis, (2) self-replication, and (3) location on the chromosomes in the nucleus (in organisms with nuclei). Does DNA (or when DNA is absent, RNA) meet these requirements?

### *Control of Enzymes*

In the next several chapters, we examine the details of protein synthesis. We will see that DNA does possess the complexity required to direct protein synthesis. Although complementarity restricts the base opposite a given base in a double helix, there are no restrictions on the sequence of bases on a given strand. Later, we will show that each sequence of three bases in DNA specifies a particular amino acid during protein synthesis. The ge-

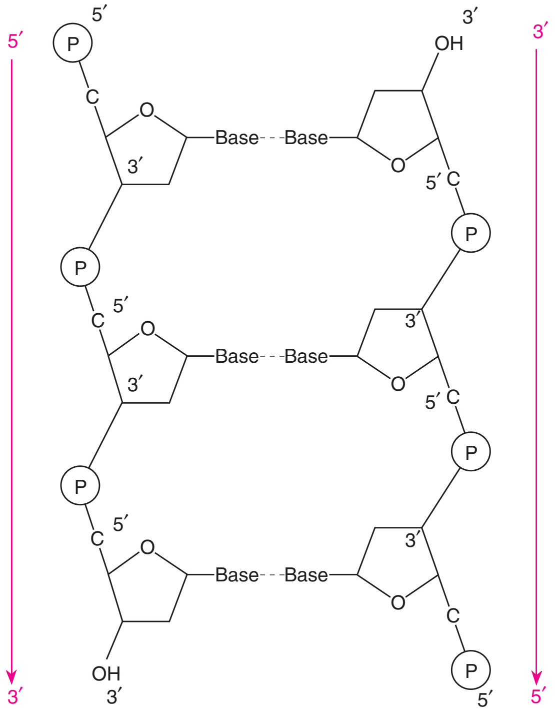


**Figure 9.15** Polarity of the DNA strands. Polarity is established by the 3′ and 5′ carbons of a given sugar. For example, moving down the left strand, the polarity is 5′ → 3′ (read as *five-prime to three-prime*). Moving down the right strand, the polarity is 3′ → 5′.

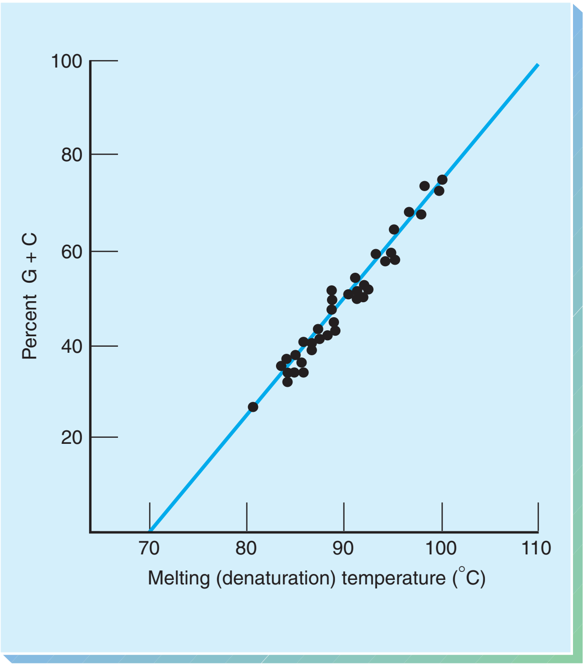


**Figure 9.16** Relationship of the number of hydrogen bonds (G-C content) and the thermal stability of DNA from different sources. (From J. Marmur and P. Doty, Jr., "Relationship of the Number of Hydrogen Bonds and the Thermal Bonds and the Thermal Stability of DNA from Different Sources," *Journal of Molecular Biology,* 5:109–112. Copyright © 1962 Academic Press LTD.)

**netic code** gives the relationship of DNA bases to the amino acids in proteins.

### *Replication*

Watson and Crick hinted in their 1953 paper how DNA might replicate. Their observation stemmed from the property of complementarity. Since the base sequence on one strand is complementary to the base sequence on the opposite strand, each strand could act as a template for a new double helix if the molecule simply "unzipped," allowing each strand to specify the sequence of bases on a new strand by complementarity (fig. 9.17). Mutability would occur due to mispairings, other errors in replication, or damage to the DNA.

### *Location*

DNA must reside in the nucleus of eukaryotes, where the genes occur on chromosomes, or in the chromosomes of prokaryotes and viruses. In both prokaryotes and eukaryotes, the majority of the cell's DNA is in the chromosomes. And all viruses contain either DNA or RNA. Thus, DNA fulfills all the requirements of a genetic material. RNA can fulfill the same requirements in RNA viruses and viroids.

## *Alternative Forms of DNA*

The form of DNA we have described so far is called **B DNA.** It is a right-handed helix: it turns in a clockwise manner when viewed down its axis. The bases are stacked almost exactly perpendicular to the main axis, with about ten base pairs per turn (34 Å; see fig. 9.13*c*). However, DNA can exist in other forms. If the water content increases to about 75%, the A form of DNA **(A DNA)** occurs. In this form, the bases tilt in regard to the axis, and there are more base pairs per turn. However, this and other known forms of DNA are relatively minor variations on the right-handed B form.

In 1979, Alexander Rich and his colleagues at MIT discovered a left-handed helix that they called **Z DNA** because its backbone formed a zigzag structure (fig. 9.18). Z DNA was found by X-ray crystallographic analysis of very small DNA molecules composed of repeating G-C sequences on one strand with the complementary C-G sequences on the other (alternating purines and pyrimidines). Z DNA looks like B DNA with each base rotated 180 degrees, resulting in a zigzag, left-handed structure (fig. 9.19). (The original configuration of the bases is referred to as the *anti* configuration; the rotated configuration is called the *syn* configuration.)

Originally, it was thought that Z DNA would not prove of interest to biologists because it required very high salt concentrations to become stable. However, it was found that Z DNA can be stabilized in physiologically normal conditions if methyl groups are added to the cytosines. Z DNA may be involved in regulating gene expression in eukaryotes. We return to this topic in chapter 16 (box 9.3).

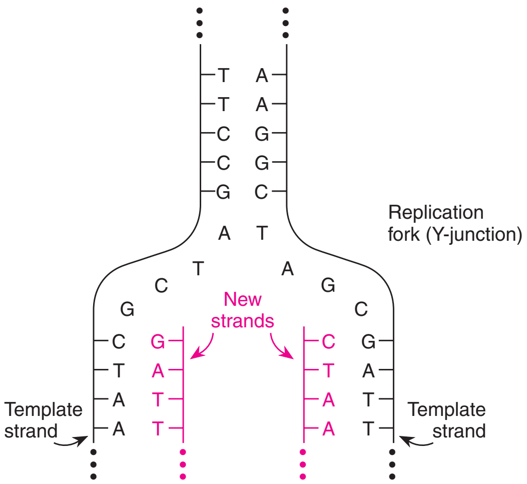


**Figure 9.17** Complementarity provides a possible mechanism for accurate DNA replication. The parent duplex opens, and each strand becomes a template for a new duplex.

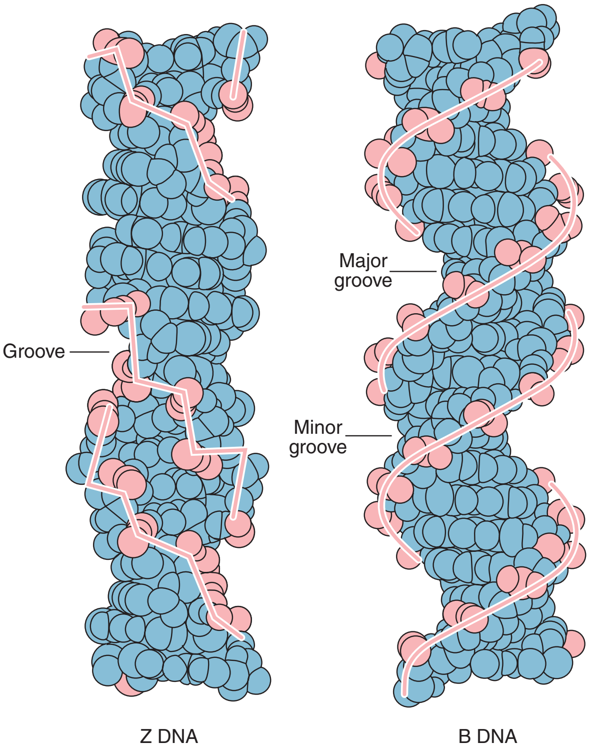


**Figure 9.18** Z (*left*) and B (*right*) DNA. The *lines* connect phosphate groups. (Reproduced with permission from the *Annual Review of Biochemistry,* Volume 53, © 1984 by Annual Reviews, Inc.)

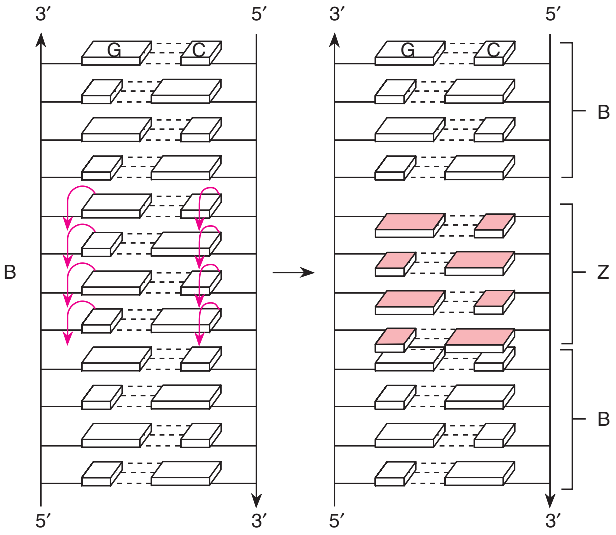


**Figure 9.19** B DNA converts to Z DNA by the rotation of bases as indicated by the curved arrows. (Reproduced with permission from *Annual Review of Biochemistry,* Volume 53, © 1984 by Annual Reviews, Inc.)

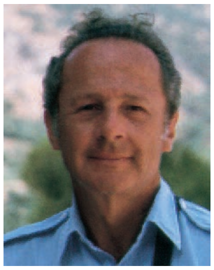
# DNA REPLICATION— THE PROCESS

In their 1953 paper, Watson and Crick hinted that the replication of the double helix could take place as the DNA unwinds, so that each strand would form a new double helix by acting as a **template** for a newly synthesized strand (see fig. 9.17). For example, when a double helix is unwound at an adenine-thymine (A-T) base pair, one unwound strand would carry A and the other would carry T. During replication, the A in the template DNA would pair with a T in a newly replicated DNA strand, giving rise to another A-T base pair. Similarly, the T in the other template strand would pair with an A in the other newly replicated strand, giving rise to another A-T base pair. Thus, one A-T base pair in one double helix would result in two A-T base pairs in two double helices. This process would repeat at every base pair in the double helix of the DNA molecule.

This mechanism is called **semiconservative** replication because, although the entire double helix is not conserved in replication, each strand is. Every daughter DNA molecule has an intact template strand and a newly replicated strand. This is not the only way that replication could occur. The alternative methods are **conservative** and **dispersive.** In conservative replication, in which the whole original double helix acts as a template for a new one, one daughter molecule would consist of the original parental DNA, and the other daughter would be totally new DNA. In dispersive replication, some parts of the original double helix are conserved, and some parts are not. Daughter molecules would consist of part template and part newly synthesized DNA. In reality, the dispersive category is the all-inclusive "other" category, including any possibility other than conservative and semiconservative replication.

## *The Meselson and Stahl Experiment*

In 1958, M. Meselson and F. Stahl reported the results of an experiment designed to determine the mode of DNA



Matthew Meselson (1930– ). (Courtesy of Dr. Matthew Meselson. Photograph by Bud Gruce.)

Franklin W. Stahl (1929– ). (Courtesy of Dr. Franklin W. Stahl.)

## BOX 9.3

### Biomedical Applications

### Multiple-Stranded DNA

Under natural conditions, single-stranded RNA and double-stranded DNA are the rule. However, under laboratory conditions, it is possible to induce a third strand of DNA to interdigitate itself into the major groove of the double helix of normal DNA in a sequence-specific fashion. That is, the third strand of DNA will not just interdigitate anywhere, but will form a stable triplex at a specific sequence (fig. 1). The rules of binding are a little less precise than normal; not all sequences are recognized, and recognition can depend on surrounding sequences. However, a thymine in the third strand will usually recognize an adenine in an adenine-thymine base pair (T•A-T), and a cytosine in the third strand will recognize a guanine in a guanine-cytosine base pair (C+G-C).

Triple-stranded nucleotide chains were first created in 1957 by three scientists at the National Institutes of Health—Alexander Rich, David Davies, and Gary Felsenfeld—while they were creating artificial nucleic acids. At the time, triple-stranded DNA seemed like a laboratory curiosity. Now it seems of interest because it may have valuable uses both experimentally and clinically. (Rich apparently had the same experience in his codiscovery of Z DNA, which at first seemed like an oddity but now is the focus of some attention—see chapter 16.) Now, researchers are able to form triplexes in naturally occurring DNA. Two applications of this technology are actively being pursued.

Both applications arise because a single strand of DNA is capable of recognizing a relatively long sequence of the double-stranded DNA in a chromosome. Thus, it is possible to selectively locate a particular genic sequence. Once the third strand locates a particular sequence on a chromosome, two things can happen. First, triplex DNA formation can prevent a particular gene from expressing itself. By the same technique, triplex DNA can also be an abortifacient, a safe method for preventing the implantation of a fetus by preventing the expression of genes under the control of the hormone progesterone.

The second use of triplex DNA is to cut DNA at a specific place by adding a cleaving compound to both ends of the third strand of DNA. Once the third strand has interdigitated, it can then break the original double helix. For example, S. Strobel and P. Dervan at the California Institute of Technology have used a chemical complex containing iron attached to both 3′ and 5′ ends of the third strand of DNA. The addition of a third chemical then initiates the cleavage reaction. The cleavage of the original duplex can be of value in modern recombinant DNA technology (see chapter 13). Whether triplex DNA will ever be of value is not certain at this time. However, it seems to have good potential for therapeutic use and to help in studying and mapping the human genome.

More recently, four-stranded DNA molecules have been found, in which double helices of certain sequences interdigitate to form four-stranded structures. These may be of importance in the formation of crossover sites or in the structures at the ends of eukaryotic chromosomes (see chapter 15).

*continued*

replication. Some historians and philosophers of science consider this the most elegant scientific experiment ever designed. Meselson and Stahl grew *E. coli* in a medium containing a heavy isotope of nitrogen, $^{15}N$. (The normal form of nitrogen is $^{14}N$.) After growing for several generations on the $^{15}N$ medium, the DNA of *E. coli* was denser. The researchers determined the density of the strands using a technique known as **density-gradient centrifugation.** In this technique, a cesium chloride (CsCl) solution is spun in an ultracentrifuge at high speed for several hours. Eventually an equilibrium arises between centrifugal force and diffusion, so that a density gradient is established in the tube with an increasing concentration of CsCl from top to bottom. If DNA (or any other substance) is added, it concentrates and forms a band in the tube at the point where its density is the same as that of the CsCl. If several types of DNA with different densities are added, they form several bands. The bands are detectable under ultraviolet light at a wavelength of 260 nm (nanometers), which nucleic acids absorb strongly.

Meselson and Stahl transferred the bacteria with heavy ($^{15}N$) DNA to a medium containing only $^{14}N$. The new DNA, replicated in the $^{14}N$ medium, was intermediate in density between light ($^{14}N$) and heavy ($^{15}N$) DNA, because the replication was semiconservative (fig. 9.20). If replication had been conservative, two bands would have appeared at the first generation of replication—an original $^{15}N$ DNA and a new $^{14}N$ double helix. And, throughout the experiment, if the method of replication had been conservative, the original DNA would have continued to show up as a $^{15}N$ band. (This, of course, did not happen.) If the method of replication had been dispersive, various multiple-banded patterns would have appeared, depending on the degree of dispersiveness. The results figure 9.20 shows are completely consistent with semiconservative replication and only with semiconservative replication.

**BOX 9.3 CONTINUED**

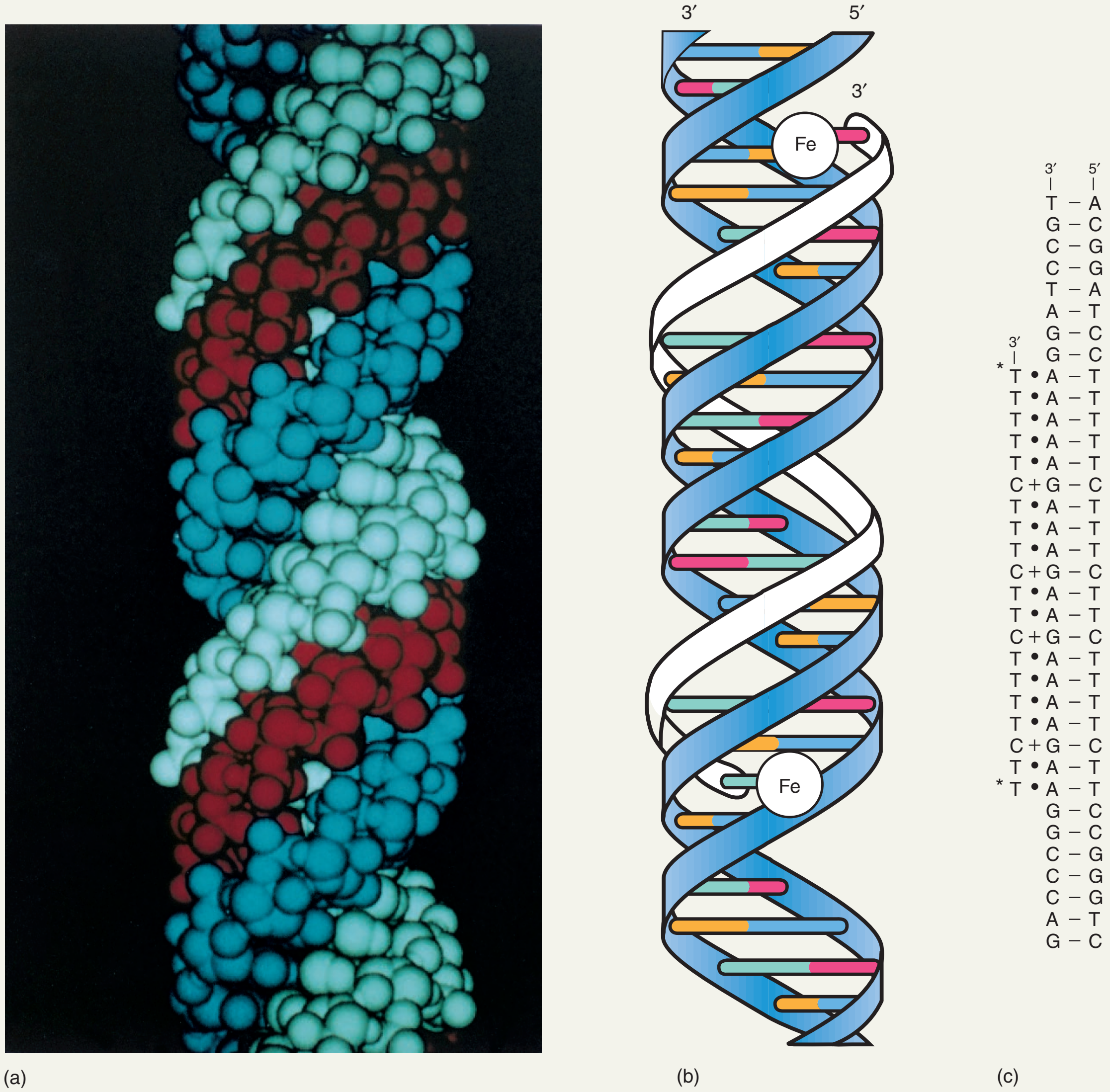


**Figure 1** A triple helix of DNA. (*a*) Space-filling model of (*b*). (*c*) Sequence of third strand on *left* and double helix on *right*. The *asterisks* in (*c*) and the *Fe* symbols in (*b*) refer to the attached iron compound. (© Courtesy Dr. Michael E. Hogan, Triplex Pharmaceutical Corporation.)

## *Autoradiographic Demonstration of DNA Replication*

In 1963, J. Cairns used **autoradiography** to verify the semiconservative method of replication photographically. This technique makes use of the fact that radioactive atoms expose photographic film. The visible silver grains on the film can then be counted to provide an estimate of the quantity of radioactive material present. Cairns grew *E. coli* bacteria in a medium containing radioactive thymine, a component of one of the DNA nucleotides. The radioactivity was in tritium ($^3$H). Cairns then carefully extracted the DNA from the bacteria and placed it on photographic emulsion for a period of time. He developed the emulsion to produce autoradiographs

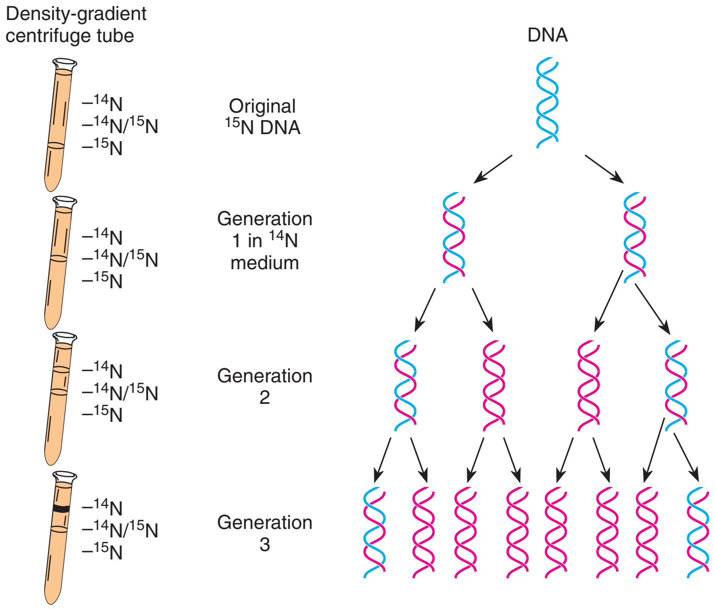


**Figure 9.20** The Meselson and Stahl experiment to determine the mode of DNA replication. The bands in the centrifuge tube are visible under ultraviolet light. The pattern of bands (*left*) comes about from semiconservative DNA replication (*right*) of $^{15}$N DNA (*blue*) replicating in a $^{14}$N medium (*red*).

that he then examined under the electron microscope (fig. 9.21). Each grain of silver represents a radioactive decay. Interpretation of this autoradiograph reveals several points. The first, known at the time, is that the *E. coli* DNA is a circle. The second point is that the DNA is replicated while maintaining the integrity of the circle. That is, the circle does not appear to break during the process of DNA replication; an intermediate **theta structure** forms (topologically similar in shape to the Greek letter theta, θ). Third, replication of the DNA seems to be occurring at one or two moving **Y-junctions** in the circle, which further supports the semiconservative mode of replication. The DNA is unwound at a given point, and replication proceeds at a Y-junction, in a semiconservative manner, in one or both directions (see fig. 9.17).

Figure 9.22 diagrams the way in which the two Y-junctions move along the circle to the final step, forming two new circles. The steps by themselves do not support either a unidirectional or a bidirectional mode of replication. That is, a theta structure will develop if either one or both Y-junctions is active in replication. But with autoradiography, it is possible to determine whether new growth is occurring in only one or in both directions.

In some cases, radioactivity was not applied to the cell until DNA replication had already begun. In these cases, the radioactive label appeared after the theta structure

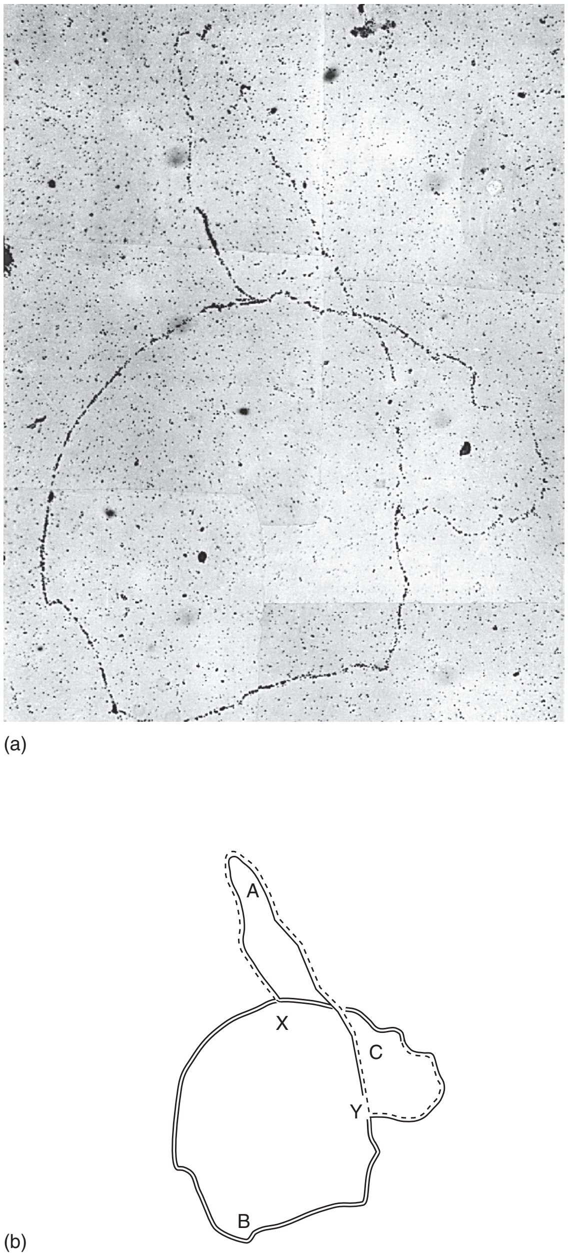


**Figure 9.21** (*a*) Autoradiograph of *E. coli* DNA during replication. (*b*) Diagram has labels on the three segments, *A, B,* and *C,* created by the existence of two forks, *X* and *Y,* in the DNA. Forks are created when the circle opens for replication. Length of the chromosome is about 1,300 μm. ([*a*] From J. Cairns, "The chromosome of *E. coli*", *Cold Spring Harbor Symposia on Quantitative Biology,* 28. Copyright © 1963 by Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Reprinted by permission.)

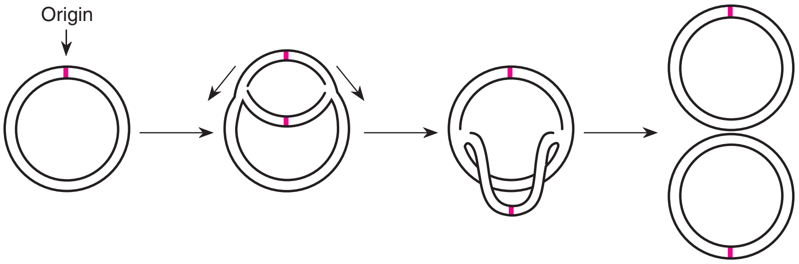


**Figure 9.22** Observable stages in the DNA replication of a circular chromosome, assuming bidirectional DNA synthesis. The intermediate figures are called theta structures.

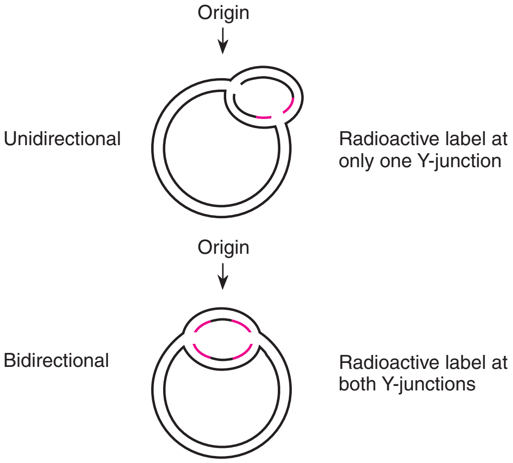

Origin
Unidirectional
Radioactive label at only one Y-junction
Origin
Bidirectional
Radioactive label at both Y-junctions

**Figure 9.23** Radioactive labels distinguish unidirectional from bidirectional DNA replication. In these hypothetical experiments, DNA replication was allowed to begin, and then a radioactive label was added. After a short period of time, the process was stopped and the autoradiographs prepared. In bidirectional replication (the actual case), the label appears at both Y-junctions.

had already begun forming. Figure 9.23 illustrates hypothetical outcomes for either unidirectional or bidirectional replication. By counting silver grains in autoradiographs, Cairns found growth to be bidirectional. Both autoradiographic and genetic analysis have subsequently verified this finding.

In eukaryotes, the DNA molecules (chromosomes) are larger than in prokaryotes and are not circular; there are also usually multiple sites for the initiation of replication. Thus, each eukaryotic chromosome is composed of many replicating units, or **replicons**—stretches of DNA with a single origin of replication. In comparison, the *E. coli* chromosome is composed of only one replicon. In eukaryotes, these replicating units form "bubbles" (or "eyes") in the DNA during replication (fig. 9.24).

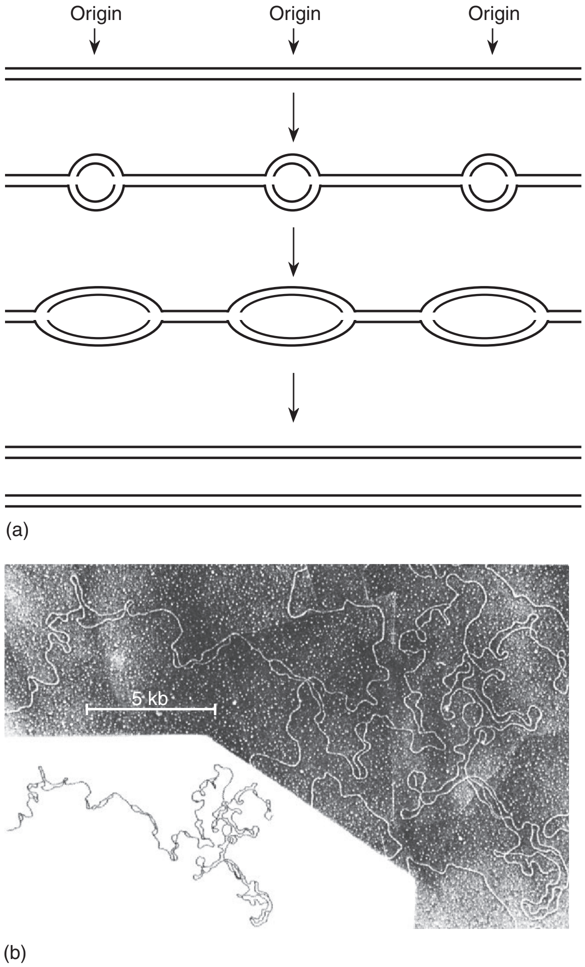
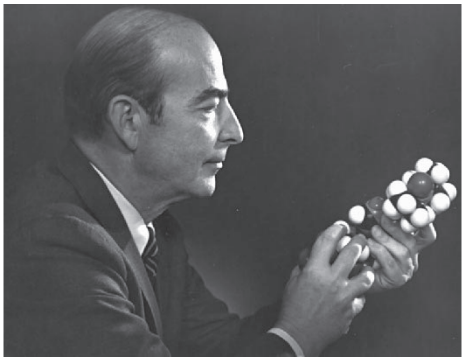


**Figure 9.24** Replication bubbles. (*a*) Formation of bubbles (eyes) in eukaryotic DNA because of multiple DNA synthesis sites of origin. (*b*) Electron micrograph (and explanatory line drawing) of replicating *Drosophila* DNA showing these bubbles. ([*b*] H. Kreigstein and D. Hogness, "Mechanism of DNA replication in *Drosophila* chromosomes: Structure of replication forks and evidence for bidirectionality," *Proceeding of the National Academy of Sciences USA,* 71 (1974):135–39. Reproduced by permission.)

## DNA REPLICATION—THE ENZYMOLOGY

Let us turn now to the details of the processes that take place during DNA replication. Like virtually all metabolic processes, DNA replication is under the control of enzymes. The evidence for the details we describe comes from physical, chemical, and biochemical studies of enzymes and nucleic acids and from the analysis of mutations that influence the replication processes. More recent techniques of recombinant DNA technology and nucleotide sequencing have allowed us to determine the nucleotide sequences of many of these key regions in DNA and RNA. We will look first at *E. coli.*

There are three major enzymes that will polymerize nucleotides into a growing strand of DNA in *E. coli.* These enzymes are **DNA polymerase** I, II, and III. DNA polymerase I, discovered by Arthur Kornberg, who subsequently won the Nobel Prize for his work, is primarily utilized in filling in small DNA segments during replication and repair processes. DNA polymerase II can serve as an alternative repair polymerase; it can also replicate DNA if the template is damaged. DNA polymerase III is the primary polymerase during normal DNA replication.



Arthur Kornberg (1918– ). (Courtesy of Dr. Arthur Kornberg. Photograph by Karsh.)

In the simplest model of DNA replication, new nucleotides would be simultaneously added, according to the rules of complementarity, on both strands of newly synthesized DNA at the replication fork as the DNA opens up. But a problem exists, created by DNA's antiparallel nature; the two strands of a DNA double helix run in opposite directions. Going in one direction on the duplex, for example, one strand is a $5' \rightarrow 3'$ strand, whereas the other is a $3' \rightarrow 5'$ strand. These directions refer to the numbering of carbon atoms across the sugar. In figure 9.25, going from the bottom of the figure to the top, the left-hand strand is a $3' \rightarrow 5'$ strand, and the right-hand strand is a $5' \rightarrow 3'$ strand. Since DNA replication involves the formation of two new antiparallel strands with the old single strands as templates, one new strand would have to be replicated in the $5' \rightarrow 3'$ direction and the other in the $3' \rightarrow 5'$ direction.

However, all the known polymerase enzymes add nucleotides in only the $5' \rightarrow 3'$ direction. That is, the polymerase catalyzes a bond between the first $5'$-$PO_4$ group of a new nucleotide and the $3'$-OH carbon of the last nucleotide in the newly synthesized strand (fig. 9.25). The polymerases cannot create the same bond with the $5'$ phosphate of a nucleotide already in the DNA and the $3'$ end of a new nucleotide. Thus, the simple model needs some revision.

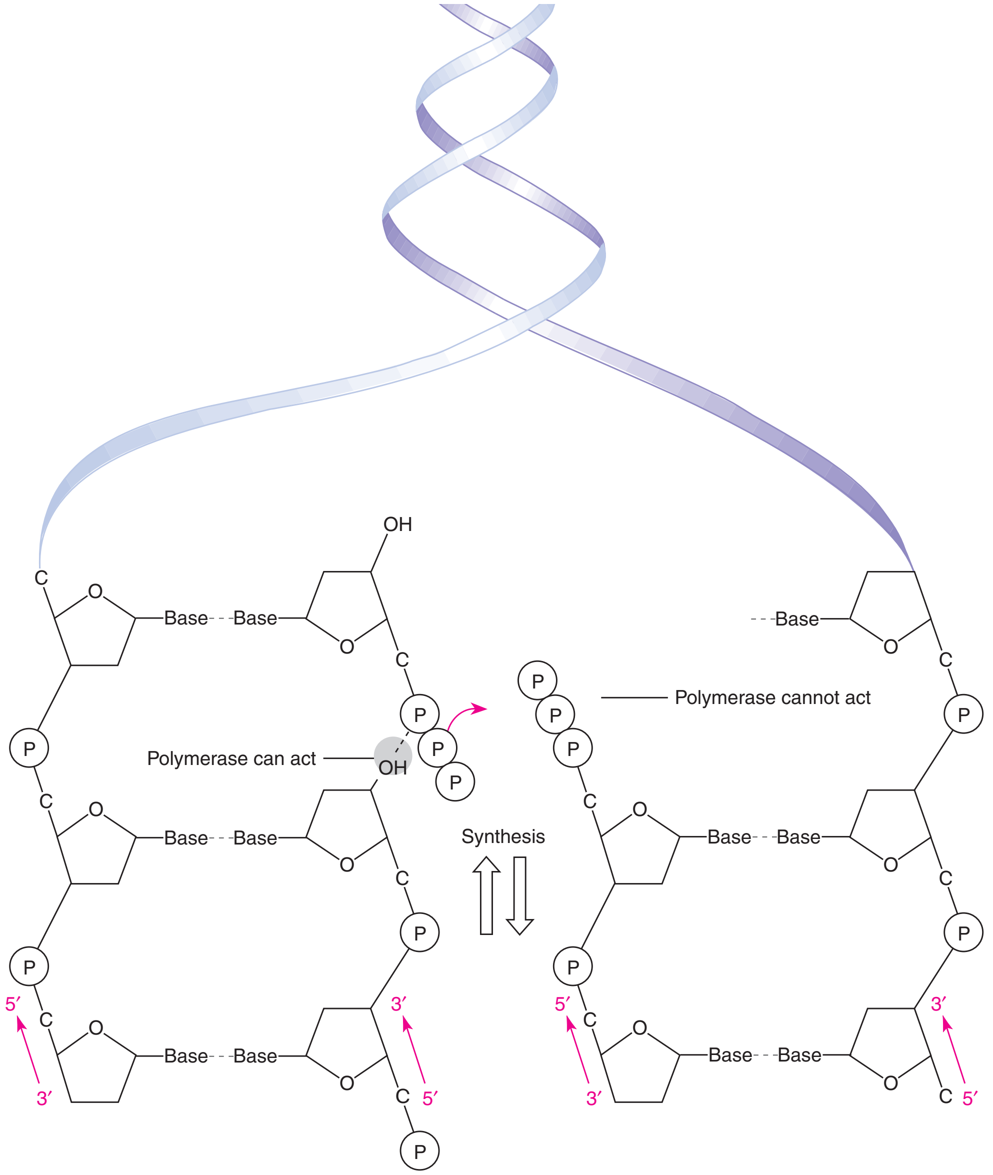
### Continuous and Discontinuous DNA Replication

Autoradiographic evidence leads us to believe that replication occurs simultaneously on both strands. **Continuous replication** is, of course, possible on the $3' \rightarrow 5'$ template strand, which begins with the necessary $3'$-OH **primer.** (Primer is double-stranded DNA—or, as we shall see, a DNA-RNA hybrid—continuing as single-stranded DNA template. The strand being synthesized has a $3'$-OH available; fig. 9.26.) A **discontinuous** form of replication takes place on the complementary strand, where it occurs in short segments, moving backward, away from the Y-junction (fig. 9.27). These short segments, called **Okazaki fragments** after R. Okazaki, who first saw them, average about 1,500 nucleotides in prokaryotes and 150 in eukaryotes. The strand synthesized continuously is referred to as the **leading strand,** and the strand synthesized discontinuously is referred to as the **lagging strand.**

Once initiated, continuous DNA replication can proceed indefinitely. DNA polymerase III on the leading-strand template has what is called high **processivity:** once it attaches, it doesn't release until the entire strand is replicated. Discontinuous replication, however, requires the repetition of four steps: primer synthesis, elongation, primer removal with gap filling, and ligation.

#### Primer Synthesis and Elongation

To synthesize Okazaki fragments, a primer must be created *de novo* (Latin: from the beginning). None of the DNA polymerases can create that primer. Instead, **primase,** an RNA polymerase coded for by the *dnaG* gene, creates the primer, ten to twelve nucleotides, at the site



OH
C
O
Base
Base
O
C
P
P
P
Polymerase can act
OH
P
C
O
Base
Base
O
C
P
P
5′
3′
C
O
Base
Base
O
3′
C
5′
P
Base
O
C
P
P
P
Polymerase cannot act
P
Synthesis
C
O
Base
Base
O
C
P
P
5′
3′
C
O
Base
Base
O
3′
C
5′

**Figure 9.25** New nucleotides can be added to DNA only during replication in the 5′ → 3′ direction.

of Okazaki fragment initiation (fig. 9.28). The result is a short RNA primer that provides the free 3′-OH group that DNA polymerase III needs in order to synthesize the Okazaki fragment. DNA polymerase III continues until it reaches the primer RNA of the previously synthesized Okazaki fragment. At that point, it stops and releases from the DNA.

All three prokaryotic polymerases not only can add new nucleotides to a growing strand in the 5′ → 3′ direction, but also can remove nucleotides in the opposite 3′ → 5′ direction. This property is referred to as *3′ → 5′ exonuclease activity.* Enzymes that degrade nucleic acids are nucleases. They are classified as **exonucleases** if they remove nucleotides from the end of a nucleotide strand or as **endonucleases** if they can break the sugar-phosphate backbone in the middle of a nucleotide strand. At first glance, exonuclease activity seems like an extremely curious property for a polymerase to have—

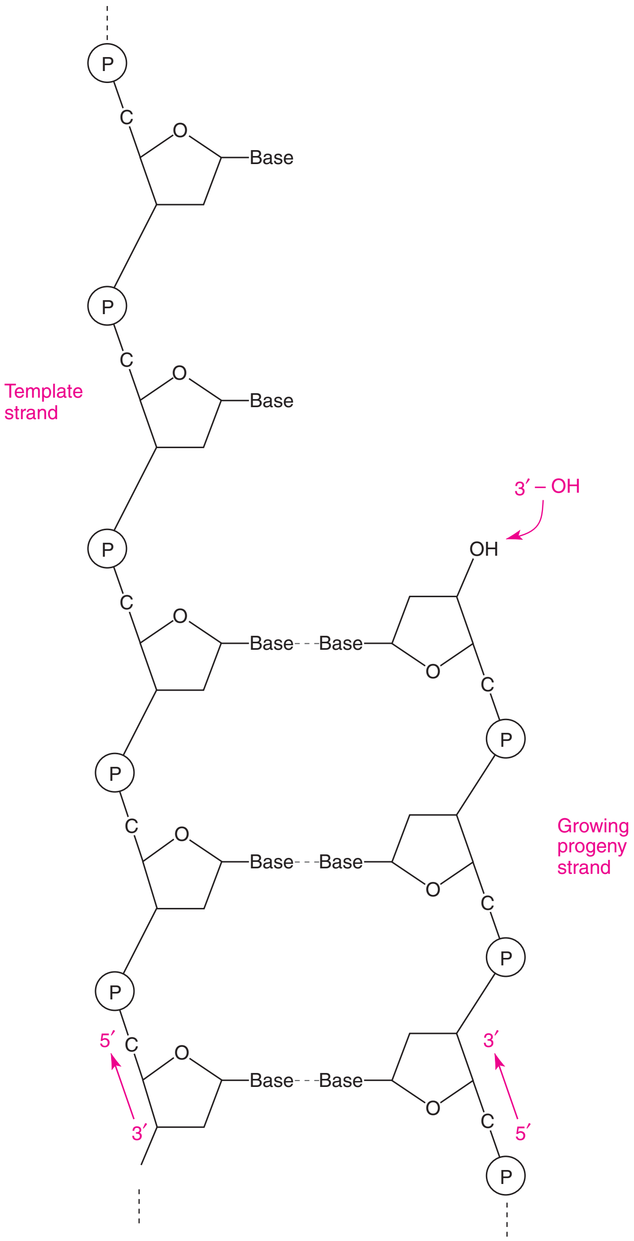


**Figure 9.26** Primer configuration for DNA replication. A 3′-OH group must be available on the nascent progeny strand opposite a continuing single-stranded template.

curious unless we think about its ability to check complementarity. If the complementarity is improper, meaning that the wrong nucleotide has been inserted, the polymerase can remove the incorrect nucleotide, put in the proper one, and continue on its way. This is known as the **proofreading** function of DNA polymerase. In addition, exonuclease activity can remove the RNA primers of Okazaki fragments.

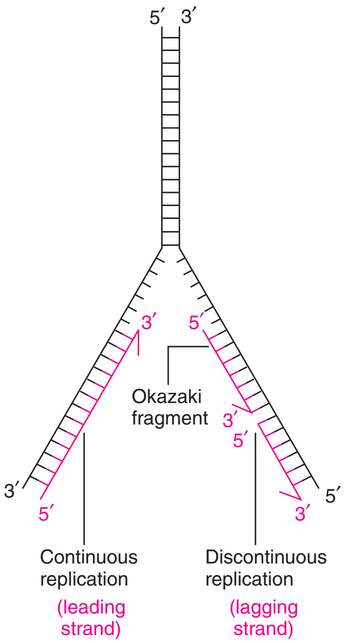


**Figure 9.27** Discontinuous model of DNA replication. Lagging-strand replication requires Okazaki fragments to form going backward, away from the Y-junction.

### *Primer Removal with Gap Filling*

DNA polymerase I is a polymerase when it adds nucleotides, one at a time, and an exonuclease when it removes nucleotides one at a time. To complete the Okazaki fragment, DNA polymerase I acts in both capacities. (DNA polymerase I mutants cannot properly connect Okazaki fragments.) DNA polymerase I completes the Okazaki fragment by removing the previous RNA primer and replacing it with DNA nucleotides (fig. 9.29). When DNA polymerase I has completed its nuclease and polymerase activity, the two previous Okazaki fragments are almost complete. All that remains is for a single phosphodiester bond to form.

### *Ligation*

DNA polymerase I cannot make the final bond to join two Okazaki fragments. The configuration needing completion is shown in figure 9.30. An enzyme, **DNA ligase,**

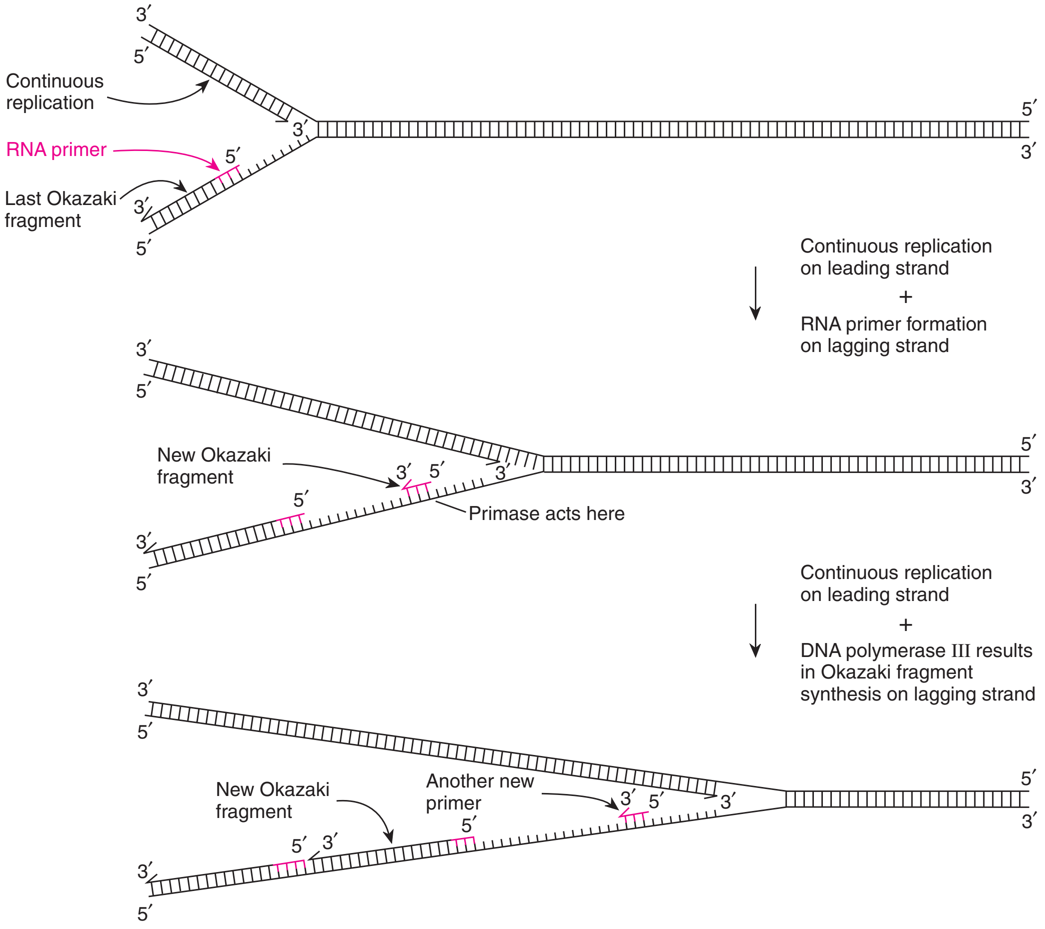
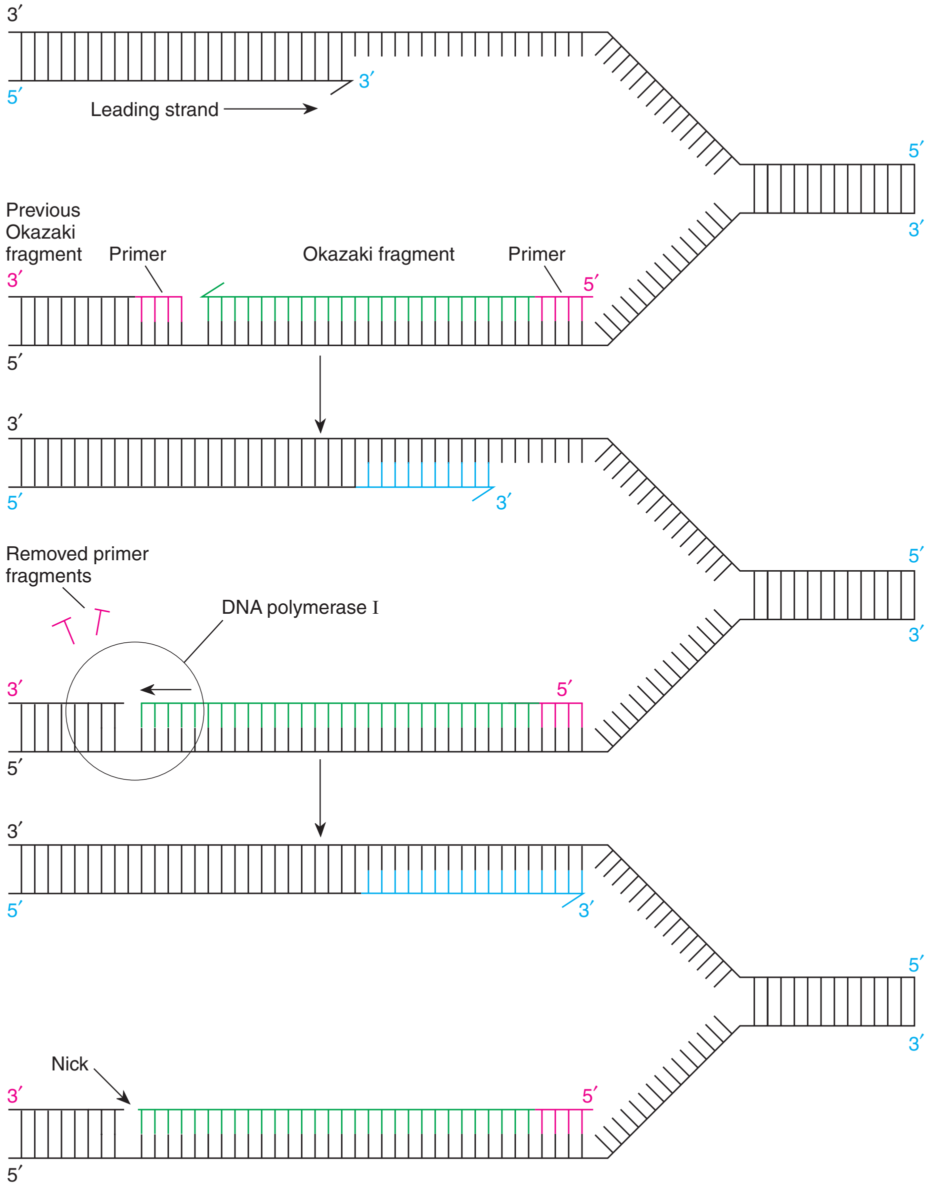

3′
5′
Continuous replication
3′
5′
3′
RNA primer
5′
Last Okazaki fragment
3′
5′
Continuous replication on leading strand
+
RNA primer formation on lagging strand
3′
5′
New Okazaki fragment
3′ 5′
3′
5′
3′
5′
3′
Primase acts here
5′
Continuous replication on leading strand
+
DNA polymerase III results in Okazaki fragment synthesis on lagging strand
3′
5′
New Okazaki fragment
Another new primer
3′ 5′
3′
5′
5′ 3′
3′
5′
5′
3′

**Figure 9.28** Primer formation and elongation create an Okazaki fragment during discontinuous DNA replication.

completes the task by making the final phosphodiester bond in an energy-requiring reaction.

A question of evolutionary interest is why RNA is used to prime DNA synthesis. Why not use DNA directly and avoid the exonuclease and resynthesis activity seen in figure 9.29? Probably, making use of RNA primers lowers the error rate of DNA replication. That is, priming is an inherently error-prone process since nucleotides are initially added without a stable primer configuration. To prevent long-term errors in the DNA, an RNA primer is put in that can later be recognized and removed. Resynthesis by polymerase I is in a much more stable primer configuration (a long primer) and thus makes very few errors.

Another question of evolutionary interest is why DNA synthesis cannot take place in the 3′ → 5′ direction. Probably, the answer has to do with proofreading and the exonuclease removal of mismatched nucleotides. When an incorrect nucleotide is found and removed, the next nucleotide brought in, in the 5′ → 3′ direction, has a triphosphate end available to provide the energy for its own incorporation (see fig. 9.25). Consider what would happen if the polymerase were capable of adding nucleotides in the opposite direction. The energy for the phosphodiester bond would be coming from the triphosphate already attached in the growing 3′ → 5′ strand (see fig. 9.25). Then, if an error in complementarity were detected and the polymerase removed the most recently added nucleotide from the 3′ → 5′ strand, the last nucleotide in the double helix would no longer have a triphosphate available to provide energy for the diester bond with the next nu-



3′
5′
3′
Leading strand
5′
3′
Previous Okazaki fragment
Primer
Okazaki fragment
Primer
3′
5′
5′
3′
5′
3′
5′
3′
Removed primer fragments
DNA polymerase I
3′
5′
5′
3′
5′
3′
5′
3′
Nick
3′
5′
5′

**Figure 9.29** The completion of an Okazaki fragment requires that DNA polymerase I replace the RNA primer base by base with DNA nucleotides. A final nick in the DNA backbone remains (*arrow*).

cleotide. Continued polymerization would thus require additional enzymatic steps to provide the energy needed for the process to continue. This could stop or slow the process down considerably. As it is, the process incorporates about four hundred nucleotides per second with an error rate of about one incorrect pairing per $10^9$ bases. (Other repair systems further improve this error rate—see chapter 12.)

## *The Origin of DNA Replication*

Each replicon (e.g., the *E. coli* chromosome, or a segment of a eukaryotic chromosome with an origin of replication) must have a region where DNA replication initiates. In *E. coli,* this region is referred to as the genetic locus *oriC;* it occurs at map location 84 minutes (see fig. 7.27). For DNA replication to begin, several steps must occur.

First, the appropriate initiation proteins must recognize the specific origin site. Then the site must be opened and stabilized. And, finally, a replication fork must be initiated in both directions, involving continuous and discontinuous DNA replication. Although most of the proteins involved are known, there are still a few gaps in our knowledge.

*OriC,* the origin of replication in *E. coli,* is about 245 base pairs long and is recognized by **initiator proteins.** These proteins, the product of the *dnaA* locus, open up the double helix. (Other DNA-binding proteins are also involved here.) The initiator proteins then take part in the attachment of DNA **helicase,** the product of the *dnaB* gene, which unwinds DNA at the Y-junction. Helicase is then responsible for recruiting (binding) the rest of the proteins that form the replication initiation complex. First is primase, which creates RNA primers. Together, the helicase and primase comprise a **primosome,** attached to the lagging-strand template. As the primosomes move along, they create RNA primers that

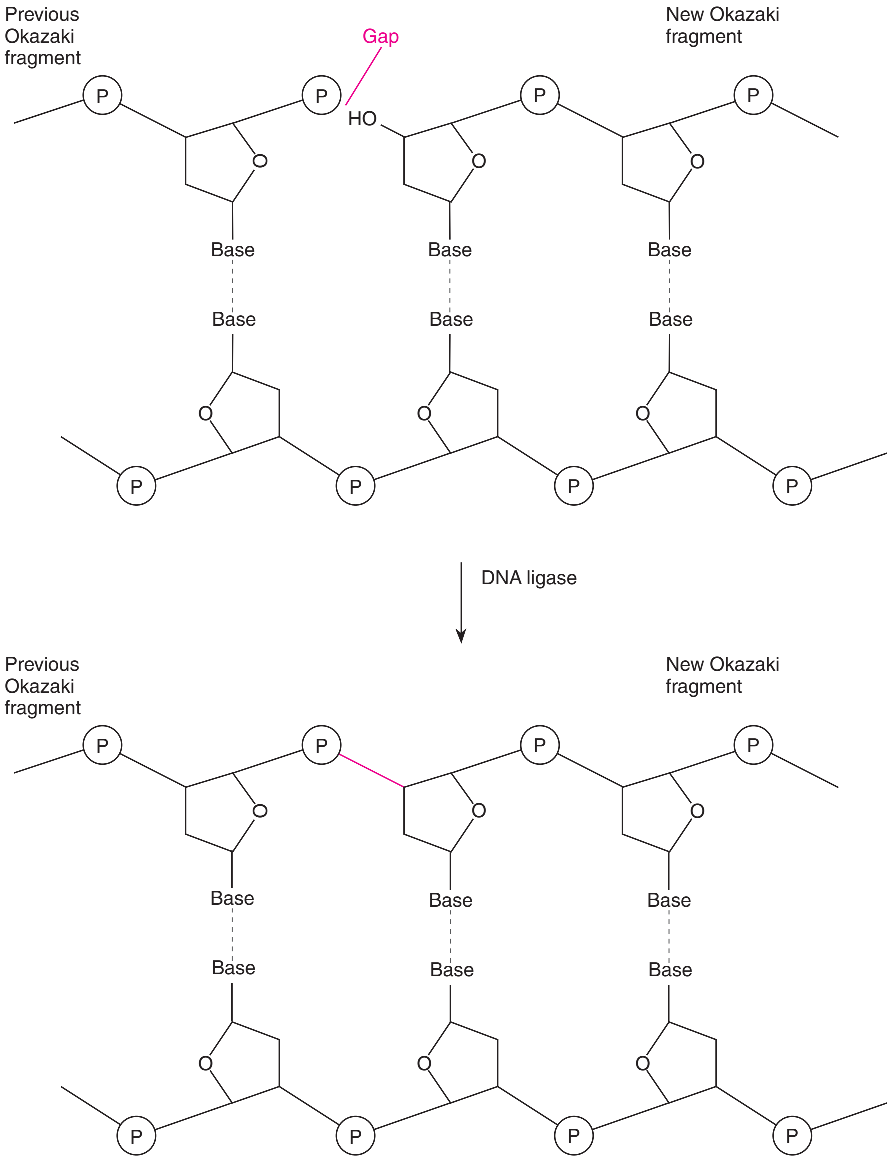


**Figure 9.30** After DNA polymerase I removes the RNA primer to complete an Okazaki fragment, a final gap remains. DNA ligase closes it.

DNA polymerase III uses to initiate leading-strand synthesis. As primers are being laid down on the lagging-strand template, Okazaki fragment synthesis begins, and Y-junction activity then proceeds as outlined earlier (see figs. 9.28, 9.29, and 9.30).

DNA polymerase III **holoenzyme** is a very large protein composed of ten subunits (table 9.4). Three of the subunits, α, ε, and θ, form the polymerization core, with both 5′ → 3′ polymerase activity and 3′ → 5′ exonuclease activity. One subunit, the β subunit, is a "processivity clamp." As a dimer (two identical copies attached head to tail), the protein forms a "doughnut" around the DNA so it can move freely on the DNA. When it is attached to the core enzyme, the polymerase is held tightly to the DNA and shows high processivity (fig. 9.31): the leading strand is usually synthesized entirely without the enzyme leaving the template (fig. 9.32). The remaining subunits are involved in processivity control and replisome formation. They allow the polymerase to move off and on the DNA of the lagging-strand template as Okazaki fragments are completed (a process known as **polymerase cycling**).

Eukaryotes have evolved at least nine DNA polymerases, named DNA polymerase α, β, γ, δ, ε, ζ, η, θ, and ι. DNA polymerase δ seems to be the major replicating enzyme in eukaryotes, forming replisomes as in *E. coli.* In eukaryotes, the polymerase α-primase complex adds the Okazaki fragment primers, first adding an RNA primer and then a short length of DNA nucleotides. Polymerase ε may be involved in repair or in normal DNA replication, as is polymerase δ. DNA polymerase γ appears to replicate mitochondrial DNA. The remaining polymerases are probably involved in DNA repair, with polymerase β being the major repair polymerase, as polymerase I is in

**Table 9.4** **Summary of the Enzymes Involved in DNA Replication in *E. coli***

| Enzyme or Protein | Genetic Locus | Function |
|---|---|---|
| DNA polymerase I | *polA* | Gap filling and primer removal |
| DNA polymerase II | *polB* | Replicating damaged templates |
| DNA polymerase III | | |
| α subunit | *dnaE* | Polymerization core; 5′ → 3′ polymerase |
| ε subunit | *dnaQ* | Polymerization core; 3′ → 5′ exonuclease |
| θ subunit | *holE* | Polymerization core |
| β subunit | *dnaN* | Processivity clamp (as a dimer) |
| τ subunit | *dnaX* | Preinitiation complex; dimerization of core |
| γ subunit | *dnaX* | Preinitiation complex; loads clamp |
| δ subunit | *holA* | Processivity core |
| δ′ subunit | *holB* | Processivity core |
| χ subunit | *holC* | Processivity core |
| ψ subunit | *holD* | Processivity core |
| Helicase | *dnaB* | Primosome; unwinds DNA |
| Primase | *dnaG* | Primosome; creates Okazaki fragment primers |
| Initiator protein | *dnaA* | Binds at origin of replication |
| DNA ligase | *lig* | Closes Okazaki fragments |
| Ssb protein | *ssb* | Binds single-stranded DNA |
| DNA topoisomerase I | *topA* | Relaxes supercoiled DNA |
| DNA topoisomerase type II (DNA Gyrase) | | |
| α subunit | *gyrA* | Relaxes supercoiled DNA; ATPase |
| β subunit | *gyrB* | Relaxes supercoiled DNA |
| Topoisomerase IV | *parE* | Unconcatenates DNA circles |
| Termination protein | *tus* | Binds at termination sites |

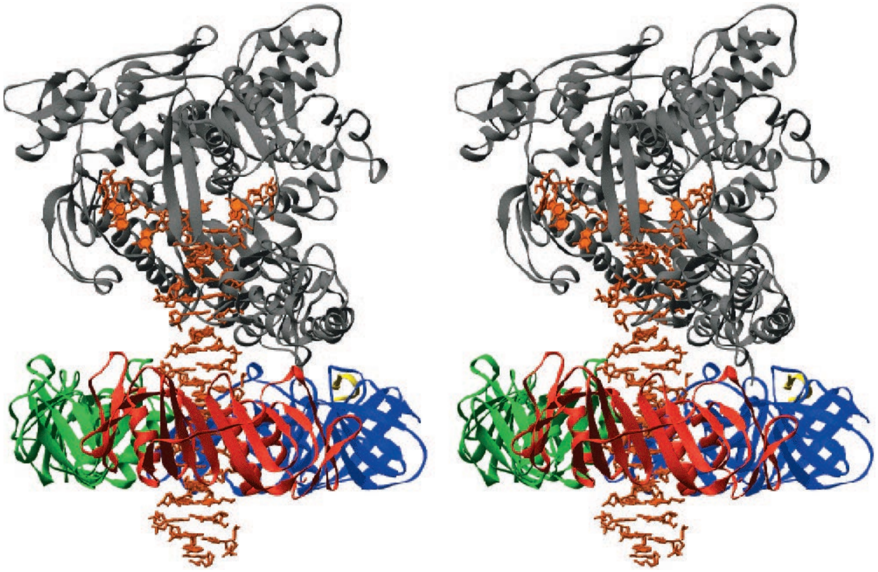


**Figure 9.31** Stereo view of sliding clamp, DNA polymerase, and DNA from bacteriophage RB69. The clamp (*red, blue, green*) surrounds the DNA (*brown*) like a doughnut. The clamp is attached to the proximal segment of the DNA polymerase (*gray*). (From Yousif Shamoo and Thomas A. Steitz, "Building a replisome from interacting pieces" in *Cell,* 99:155–166, October 15. Reprinted by permission of *Cell.*)

*E. coli.* Several of the polymerases most likely both replicate and repair DNA.

Eukaryotes also have a clamp-loader complex, called replication factor C, and a six-unit clamp called the proliferating cell nuclear antigen. The RNA primers are removed during Okazaki fragment completion (maturation) by mechanisms similar to those in prokaryotes. In eukaryotes, RNAase enzymes remove the RNA primers in Okazaki fragments; a repair polymerase fills gaps; and a DNA ligase forms the final seal. Helicases, topoisomerases, and single-strand binding proteins play roles similar to those they play in prokaryotes. The completion of the replication of linear eukaryotic chromosomes involves the formation of specialized structures at the tips of the chromosomes, which we discuss in chapter 15. Thus, all of the enzymatic processes are generally the same in prokaryotes and eukaryotes. DNA replication developed in prokaryotes and was refined as prokaryotes evolved into eukaryotes.

T. Steitz and his colleagues have done much X-ray crystallography work that has given us an excellent look at the structure of a polymerase. (Most work has actually been done on a fragment of DNA polymerase I called the **Klenow fragment.**) The enzyme is shaped like a cupped right hand with enzymatic activity taking place in two places, separated by a distance of about two to three nucleotides (fig. 9.33). It is proposed that when the polymerization site senses a mismatch, the DNA is moved so that the 3′ end enters the exonuclease site, where the incorrect nucleotide residue is then cleaved. Polymerization then continues. There may be a general mode of polymerase action among diverse polymerases.

The replication of the *E. coli* chromosome may be controlled by the methylation state of several sequences within *oriC.* As we discuss in chapter 13, certain enzymes add methyl groups to specific DNA bases, and the presence or absence of these methyl groups can serve as signals to other enzymes.

## Events at the Y-Junction

We now have the image of DNA replication proceeding as a primosome moves along the lagging-strand template, opening up the DNA (helicase activity), and creating RNA primers (primase activity) for Okazaki fragments. One DNA polymerase III moves along the leading-strand template, generating the leading strand by continuous DNA replication, whereas a second DNA polymerase III moves backward, away from the Y-junction, creating Okazaki fragments. **Single-strand binding proteins** (ssb proteins) keep single-stranded DNA stabilized (open) during

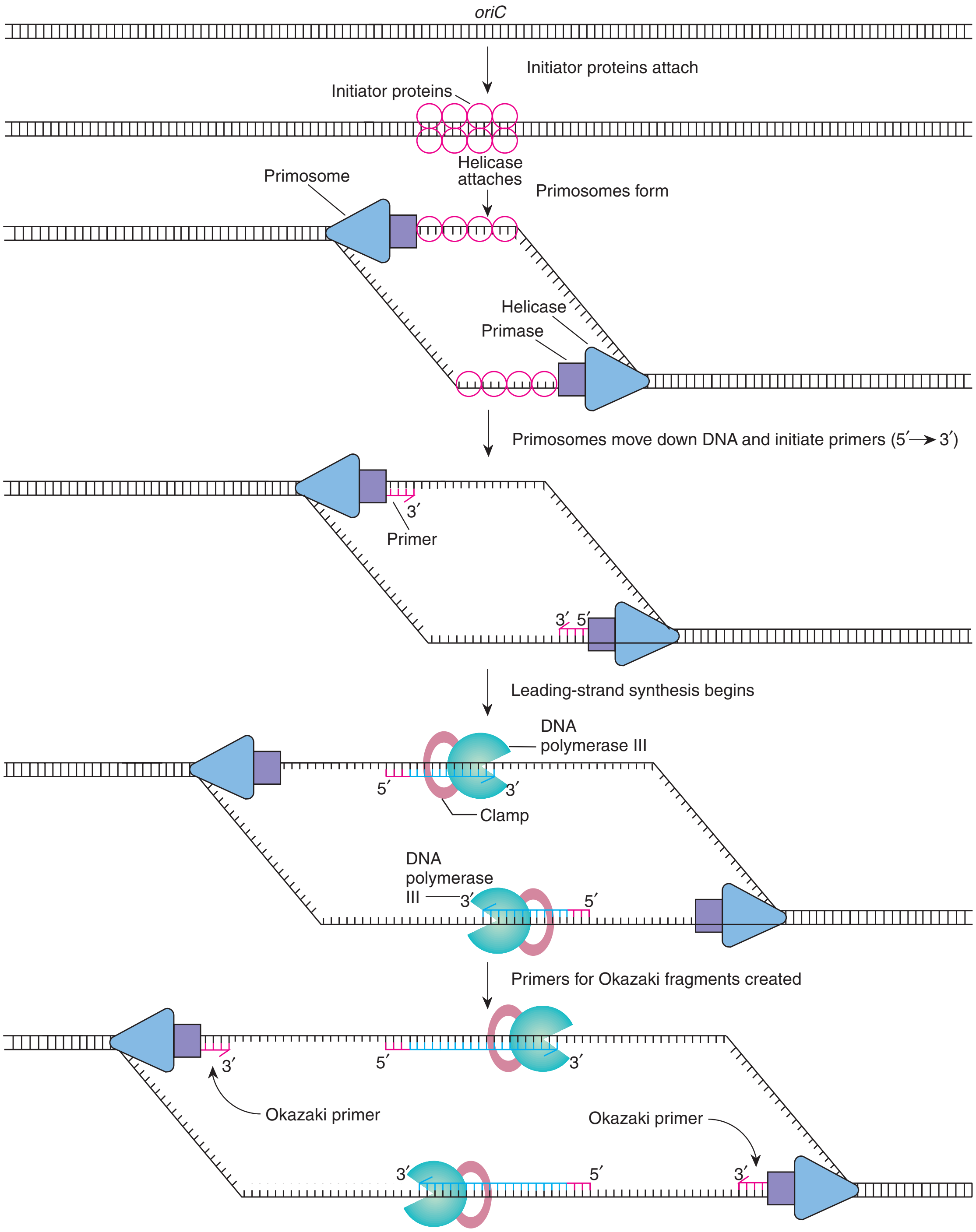

oriC
Initiator proteins attach
Initiator proteins
Primosome
Helicase attaches
Primosomes form
Helicase
Primase
Primosomes move down DNA and initiate primers (5′⟶3′)
3′
Primer
3′ 5′
Leading-strand synthesis begins
DNA polymerase III
5′
3′
Clamp
DNA polymerase III
3′
5′
Primers for Okazaki fragments created
3′
5′
3′
Okazaki primer
Okazaki primer
3′
5′
3′

**Figure 9.32** Events at the origin of DNA replication in *E. coli.* The DNA opens up at *oriC* to create two moving Y-junctions. Initiator proteins attach and then bind helicase. The helicase then binds primase, forming a primosome. After the primer forms and two copies of DNA polymerase III are bound, the polymerization process begins.

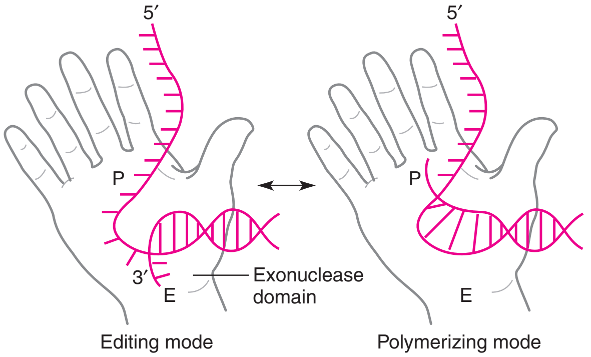


**Figure 9.33** DNA polymerase (P) and exonuclease (E) activities of the Klenow fragment of DNA polymerase I in *E. coli.* On the *right,* 5′ → 3′ polymerization is occurring. On the *left,* the 3′ end of the nascent strand has been backed up into the exonuclease site, presumably when a mismatch was detected. (With permission from the *Annual Review of Biochemistry,* Volume 63. ©1994 by Annual Reviews. www.AnnualReviews.org.)

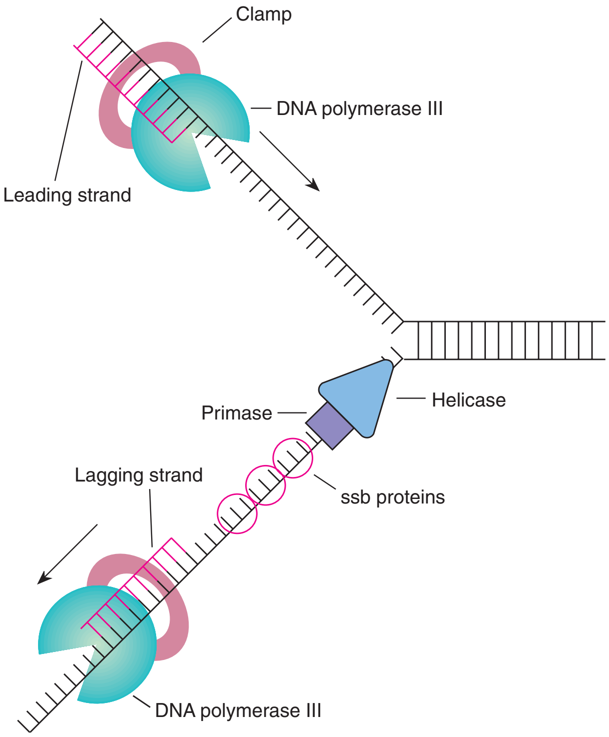


**Figure 9.34** Schematic drawing of DNA replication at a Y-junction. Two copies of DNA polymerase III, ssb proteins, and a primosome (helicase + primase) are present.

this process, and DNA polymerase I and ligase connect Okazaki fragments (fig. 9.34).

This simple picture is slightly complicated by the fact that the lagging- and leading-strand synthesis is coordinated. B. Alberts suggested an explanation: the **replisome** model, in which both copies of DNA polymerase III are attached to each other and work in concert with the primosome at the Y-junction (fig. 9.35). According to this model, a single replisome, consisting of two copies of DNA polymerase III, a helicase, and a primase, moves along the DNA. The leading-strand template is immediately fed to a polymerase, whereas the lagging-strand template is not acted on by the polymerase until an RNA primer has been placed on the strand, meaning that a long (fifteen-hundred base) single strand has been opened up (fig. 9.35*a*).

As the replisome moves along, another single-stranded length of the lagging-strand template forms. At about the time that the Okazaki fragment is completed, a new RNA primer has been created (fig. 9.35*b*). The Okazaki fragment is released (fig. 9.35*c*), and a new Okazaki fragment is begun (polymerase cycling), starting with the latest primer (fig. 9.35*d*). This takes the replisome back to the same configuration as in figure 9.35*a,* but one Okazaki fragment farther along.

Figure 9.36 gives us a closer look at the details of the Y-junction at the moment of polymerase cycling. Primase, which is not highly processive, must be in touch with an ssb protein to stay attached to the DNA when forming a primer. At the appropriate moment, after the primer is formed, the clamp loader contacts the ssb, dislodging the primase. The clamp loader also loads a sliding clamp, which then recruits (attaches to) the polymerase that is creating the lagging strand. The polymerase then continues, creating the Okazaki fragment. The primase can later attach at a new point on the lagging-strand template to create the next primer.

## Supercoiling

The simplicity and elegance of the DNA molecule masks an inevitable problem: coiling. Since the DNA molecule is made from two strands that wrap about each other, certain operations, such as DNA replication and its termination, face topological difficulties. Up to this point, we have seen the circular *E. coli* chromosome in its “relaxed” state (e.g., figs. 9.21 and 9.22). However, certain enzymes in the cell cause DNA to become overcoiled (positively **supercoiled**) or undercoiled (negatively supercoiled). Positive supercoiling comes about in two ways: either the DNA takes too many turns in a given length, or the molecule wraps around itself (fig. 9.37).

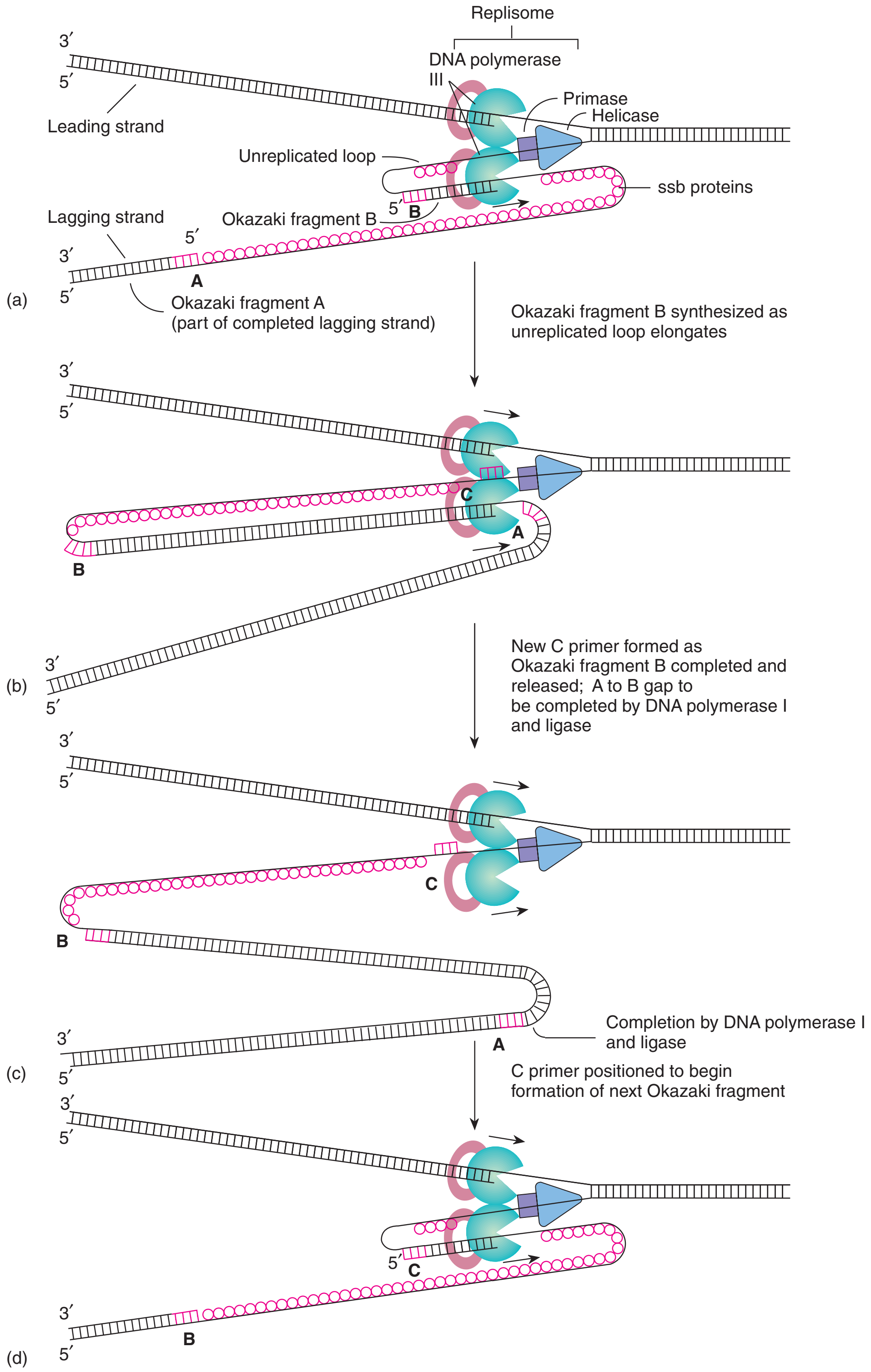

Replisome
DNA polymerase III
Primase
Helicase
3′
5′
Leading strand
Unreplicated loop
ssb proteins
Lagging strand
Okazaki fragment B
5′
B
A
Okazaki fragment A
(part of completed lagging strand)
(a)
Okazaki fragment B synthesized as unreplicated loop elongates
C
A
B
(b)
New C primer formed as Okazaki fragment B completed and released; A to B gap to be completed by DNA polymerase I and ligase
C
B
A
(c)
Completion by DNA polymerase I and ligase
C primer positioned to begin formation of next Okazaki fragment
5′
C
B
(d)

**Figure 9.35** The replisome, which consists of two DNA polymerase III holoenzymes and a primosome (helicase + primase), coordinates replication at the Y-junction. Parts *b–d* show "polymerase cycling," in which the polymerase on the lagging-strand template releases a completed Okazaki fragment and then begins the next one.

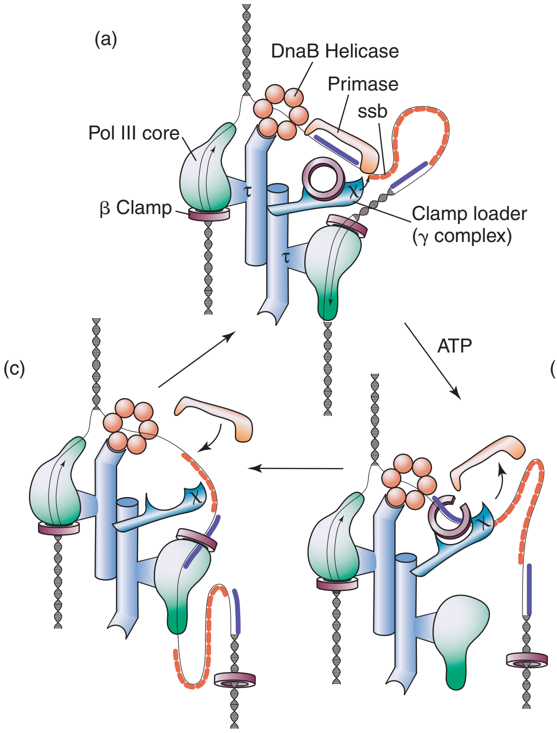


**Figure 9.36** A close-up view of the Y-junction during polymerase cycling. The two polymerases (pol III core) are held together by τ subunits. Also pictured are the sliding clamp (β Clamp), clamp loader (γ complex), primase, helicase, and ssb proteins. In (*a*), the primase has just finished creating a primer. The χ subunit of the clamp loader contacts the ssb protein that is touching the primase; the primase is then dislodged (*b*). The clamp is loaded at the new primer and the polymerase on the lagging strand is cycled to the clamp to begin the next Okazaki segment (*c*). (Reprinted from *Cell,* Vol. 96, Yuzhakov et al., "Trading Places on DNA-a Three Point Switch Underlies Primer Handoff from Primase to the Replicative DNA Polymers," pp. 153–163, Copyright © 1999, with permission from Elsevier Science.)

Positive supercoiling occurs when the circular duplex winds about itself in the same direction as the helix twists (right-handed), whereas negative supercoiling comes about when the duplex winds about itself in the opposite direction as the helix twists (left-handed). The former increases the number of turns of one helix around the other (the **linkage number,** L), whereas the latter decreases it. The three forms of DNA in figure 9.37 all have the same sequence, yet they differ in linkage number. Accordingly, they are referred to as topological isomers **(topoisomers).** The enzymes that create or alleviate these states are called **topoisomerases.**

Topoisomerases affect supercoiling by either of two methods. Type I topoisomerases break one strand of a double helix and, while binding the broken ends, pass the other strand through the break. The break is then sealed (fig. 9.38). Type II topoisomerases (e.g., **DNA gyrase** in *E. coli*) do the same sort of thing, only instead of breaking one strand of a double helix, they break both and pass another double helix through the temporary gap. Four topoisomerases are active in *E. coli,* with somewhat confusing nomenclature: topoisomerases I and III are type I; topoisomerases II and IV are type II.

As DNA replication proceeds, positive supercoiling builds up ahead of the Y-junction. This is eliminated by topoisomerases that either create negative supercoiling ahead of the Y-junction in preparation for replication or alleviate positive supercoiling after it has been created.

## Termination of Replication

The termination of the replication of a circular chromosome presents no major topological problems. At the end of the theta-structure replication (see fig. 9.22), both Y-junctions have proceeded around the molecule. The region of termination on the *E. coli* chromosome, the terminus region, is 180 degrees from *oriC* on the circular chromosome, between minutes 28 and 36. There are six terminator sites (*Ter*); three arrest the Y-junction from the left, and three arrest the one from the right when bound by a termination protein, the protein product of the *tus* gene. (*Tus* stands for terminus utilization substance; each *Ter* site is about twenty base pairs.) One interesting aspect of the termination of *E. coli* DNA replication is that the cells are viable even if the whole terminator region is deleted. There are fewer viable cells and some growth problems, but in general, *E. coli* can successfully terminate DNA replication even without formal termination sites. A topoisomerase, topoisomerase IV, then releases the two circles, and DNA polymerase I and ligase close them up (fig. 9.39).

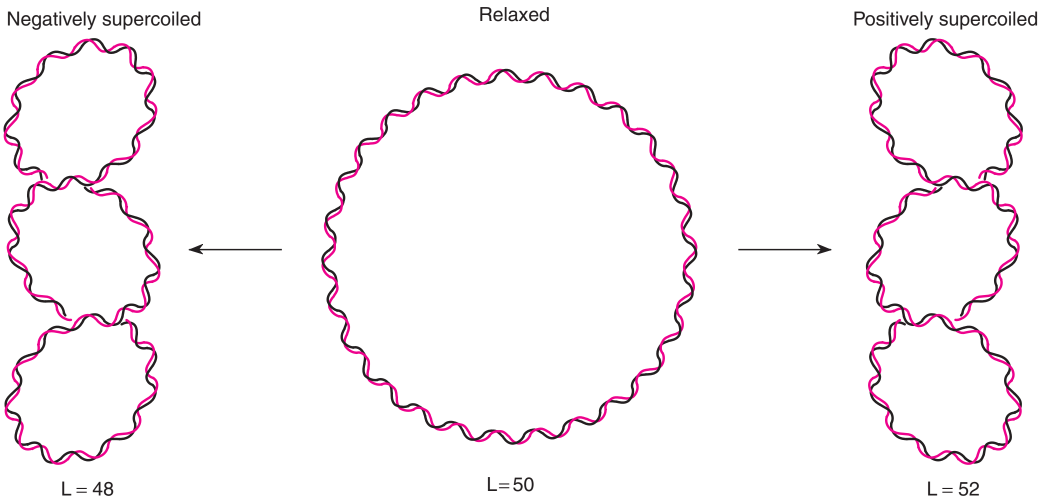

Negatively supercoiled
Relaxed
Positively supercoiled
L = 48
L = 50
L = 52

**Figure 9.37** Positive and negative supercoils. Enzymes called topoisomerases can take relaxed DNA (*center*) and add negative (*left*) or positive (*right*) supercoils. *L* is the linkage number.

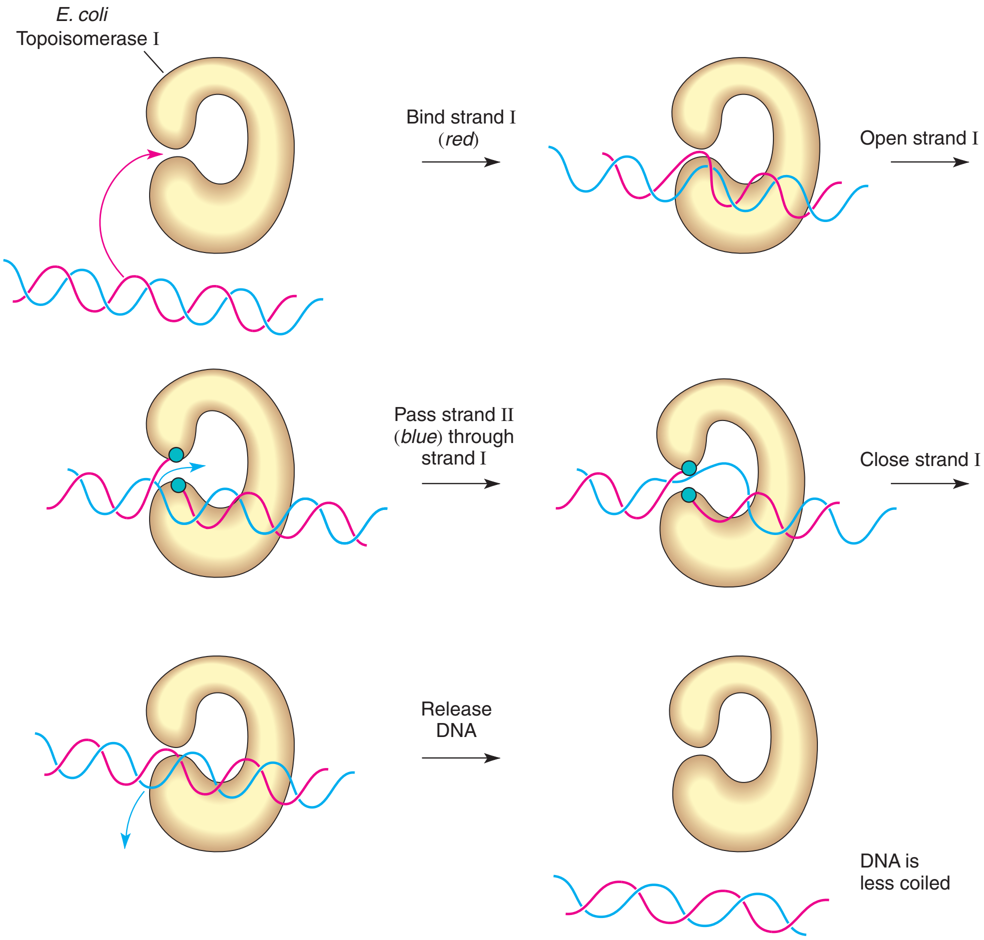

*E. coli*
Topoisomerase I
Bind strand I
(*red*)
Open strand I
Pass strand II
(*blue*) through
strand I
Close strand I
Release
DNA
DNA is
less coiled

**Figure 9.38** Topoisomerase I can reduce DNA coiling by breaking one strand of the double helix and passing the other strand through it.

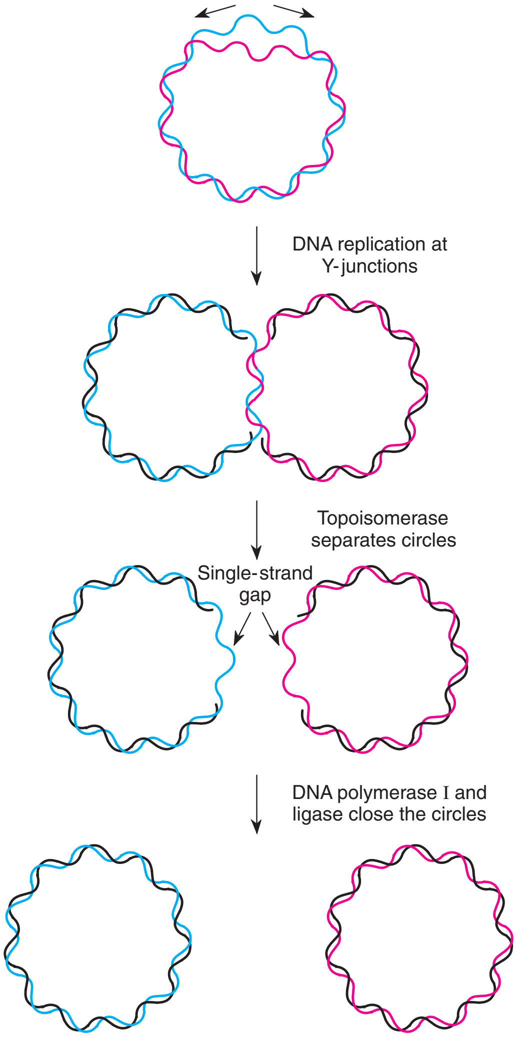

DNA replication at Y-junctions
Topoisomerase separates circles
Single-strand gap
DNA polymerase I and ligase close the circles

**Figure 9.39** The replication of circular DNA terminates when topoisomerase separates the circles and DNA polymerase I and ligase close the gaps in each circle.

### *DNA Partitioning in* E. coli

In chapter 3, we discussed processes that partition eukaryotic chromosomes between daughter cells during mitosis and meiosis. Until very recently, geneticists believed that the partitioning of the *E. coli* chromosome was a passive process, unlike that in eukaryotes. Now, however, we know that more complexity is involved in *E. coli* DNA partitioning. When DNA replication begins, the newly replicated origins of replication are segregated to opposite ends of the bacterial cell, acting as centromeres do. A ring of proteins, the products of the *FtsZ* gene, form a ring at the middle of the cell and begin to create the septum that will divide the cell into two. The full complexity involved in *E. coli* chromosomal partitioning should be uncovered in the near future.

## REPLICATION STRUCTURES

The *E. coli* model of DNA replication that we have presented here is by way of the intermediate theta-structure (see fig. 9.22). Two other modes of replication occur in circular chromosomes: rolling-circle and D-loop.

### *Rolling-Circle Model*

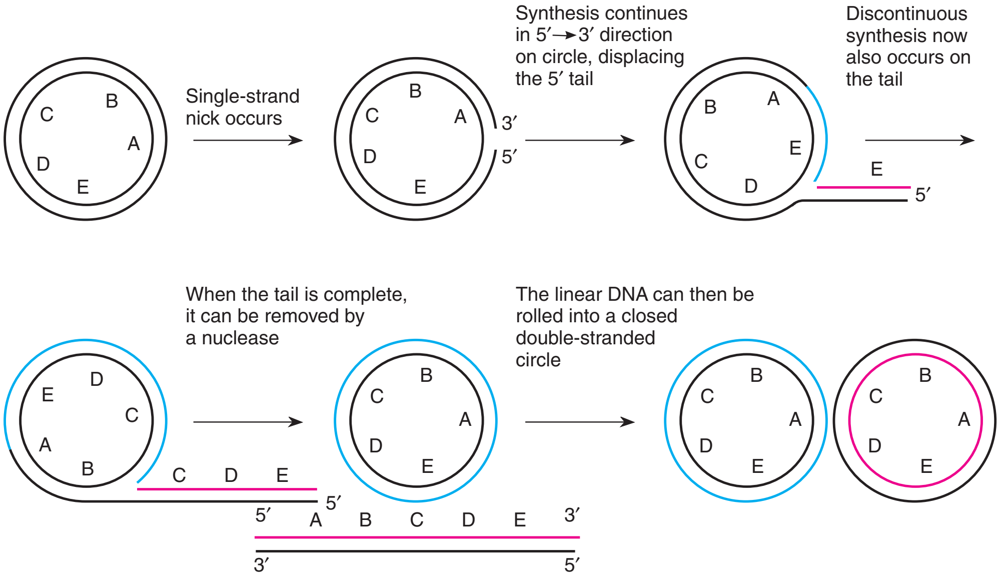
In the **rolling-circle** mode of replication, a nick (a break in one of the phosphodiester bonds) is made in one of the strands of the circular DNA, resulting in replication of a circle and a tail (fig. 9.40). This form of replication occurs in the F plasmid or *E. coli* Hfr chromosome during conjugation (see chapter 7). The $F^+$ or Hfr cell retains the circular daughter while passing the linear tail into the $F^-$ cell, where replication of the tail takes place. Several phages also use this method, filling their heads (protein coats) with linear DNA replicated from a circular parent molecule.

### *D-Loop Model*

Chloroplasts and mitochondria (in eukaryotic cells) have their own circular DNA molecules (see chapter 17) that appear to replicate by a slightly different mechanism. The origin of replication is at a different point on each of the two parental template strands. Replication begins on one strand, displacing the other while forming a displacement loop or **D-loop** structure (fig. 9.41). Replication continues until the process passes the origin of replication on the other strand. Replication then initiates on the second strand, in the opposite direction. Normal Y-junction replication, as described earlier, also occurs in mitochondrial DNA under some growth conditions.

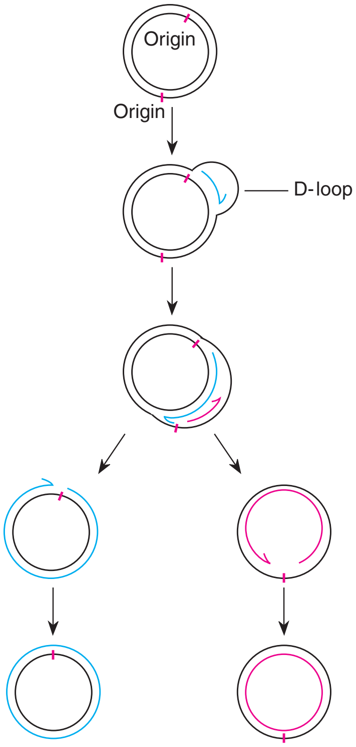
## EUKARYOTIC DNA REPLICATION

As we saw earlier, linear eukaryotic chromosomes usually have multiple origins of replication, resulting in figures referred to as "bubbles" or "eyes" (see fig. 9.24). Multiple origins allow eukaryotes to replicate their larger quantities of DNA in a relatively short time, even though eukaryotic DNA replication is considerably slowed by the presence of histone proteins associated



Single-strand nick occurs
Synthesis continues in 5′→3′ direction on circle, displacing the 5′ tail
Discontinuous synthesis now also occurs on the tail
When the tail is complete, it can be removed by a nuclease
The linear DNA can then be rolled into a closed double-stranded circle
A
B
C
D
E
3′
5′

**Figure 9.40** Rolling-circle model of DNA replication. The letters *A–E* provide landmarks on the chromosomes.



Origin
Origin
D-loop

**Figure 9.41** D-loops form during mitochondrial and chloroplast DNA replication because the origins of replication are at different places on the two strands of the double helix.

with the DNA to form chromatin (see chapter 15). For example, the *E. coli* replication fork moves through about twenty-five thousand base pairs per minute, whereas the eukaryotic Y-junction moves through only about two thousand base pairs per minute. The number of replicons in eukaryotes varies from about five hundred in yeast to as many as sixty thousand in a diploid mammalian cell.

In budding yeast, a lower eukaryote that is often used as a model organism, DNA replication initiates at sites called **autonomously replicating sequences (ARS).** Each consists of a specific 11-base-pair sequence plus two or three additional short DNA sequences encompassing 100–200 base pairs. Six proteins form a complex that binds to this sequence, referred to as the **origin recognition complex (ORC).** These proteins seem to be bound all the time, and thus additional proteins are needed to initiate DNA replication. Some of these additional proteins are cyclin-dependent kinases, proteins involved in the control of the cell cycle (chapter 3). This makes sense because in eukaryotes, DNA replication can take place only once during the cell cycle, during the S phase. Thus, the initiation of DNA replication must be tightly controlled to avoid multiple replication of some or all replicons.

## SUMMARY

**STUDY OBJECTIVE 1:** To understand the properties that a genetic material must have 205–211.

A genetic material must be able to control the phenotype of a cell or organism (i.e., to direct protein synthesis), it must be able to replicate, and it must be located in the chromosomes. Avery and his colleagues demonstrated that DNA was the genetic material when they showed that the transforming agent was DNA. Griffith had originally demonstrated transformation of *Streptococcus* bacteria in mice. Hershey and Chase demonstrated that the DNA of bacteriophage T2 entered the bacterial cell. Fraenkel-Conrat demonstrated that in viruses without DNA (RNA viruses), such as tobacco mosaic virus, the RNA acted as the genetic material. Thus, by 1953, the evidence strongly suggested nucleic acids (DNA or, in its absence, RNA) as the genetic material.

**STUDY OBJECTIVE 2:** To examine the structure of DNA, the genetic material 211–224.

Chargaff showed a 1:1 relationship of adenine (A) to thymine (T) and cytosine (C) to guanine (G) in DNA. Wilkins, Franklin, and their colleagues showed, by X-ray crystallography, that DNA was a helix of specific dimensions. Following these lines of evidence, Watson and Crick in 1953 suggested the double-helical model of the structure of DNA. In their model, DNA is made up of two strands, running in opposite directions, with sugar-phosphate backbones and bases facing inward. Bases from the two strands form hydrogen bonds with each other with the restriction that only A and T or G and C can pair. This explains the quantitative relationships that Chargaff found among the bases. Melting temperatures of DNA also support this structural hypothesis because DNAs with higher G-C contents have higher melting, or denaturation, temperatures; G-C base pairs have three hydrogen bonds versus only two in an A-T base pair. The Watson-Crick DNA model represents the B form. DNA can exist in other forms, including the Z form, a left-handed double helix that may be important in controlling eukaryotic gene expression.

**STUDY OBJECTIVE 3:** To investigate the way in which DNA replicates 220–239.

DNA replicates by unwinding of the double helix, with each strand subsequently acting as a template for a new strand. This works because of complementarity—only A-T, T-A, G-C, or C-G base pairs form stable hydrogen bonds within the structural constraints of the model. This model of replication is *semiconservative.* Meselson and Stahl confirmed it in an experiment with heavy nitrogen. Autoradiographs of replicating DNA showed that replication proceeds bidirectionally from a point of origin. Prokaryotic chromosomes are circular, with a single initiation point of replication. Eukaryotic DNA is linear, with multiple initiation points of replication.

DNA polymerase enzymes add nucleotides only in the $5' \rightarrow 3'$ direction. Replication proceeds in small segments, working backward from the Y-junction on the $5' \rightarrow 3'$ template strand. Presumably, the $5' \rightarrow 3'$ restriction has to do with the proofreading DNA polymerases do to correct errors in complementarity. Polymerase III is the active replicating enzyme, and polymerase I is involved in DNA repair. Many other enzymes help create the Okazaki fragments, unwind DNA, and release the DNA from supercoiling. Prokaryotic and eukaryotic systems follow similar steps.

## SOLVED PROBLEMS

**PROBLEM 1:** What evidence led to the idea that DNA was the genetic material?

***Answer:*** Avery and his colleagues (MacLeod and McCarty) performed experiments showing that DNA was the transforming agent, and they are thus generally given credit for formalizing the notion that DNA, not protein, is the genetic material. Chargaff, Hershey and Chase, Fraenkel-Conrat, and several others also helped shape the general view. At the time that Watson and Crick published their model, the scientific community knew that DNA was the genetic material but didn't know its structure.

**PROBLEM 2:** How does DNA fulfill the requirements of a genetic material?

***Answer:*** DNA is located in chromosomes, has a structure that is easily and accurately replicated, and has the sequence complexity to code for the fifty thousand or more genes that a eukaryotic organism has.

**PROBLEM 3:** What enzymes are involved in DNA replication in *E. coli?*

***Answer:*** A replisome, consisting of a primosome (a primase and a helicase) and two polymerase III holoenzymes, forms at a Y-junction on DNA. One polymerase acts processively, synthesizing the leading strand, while the other forms Okazaki fragments initiated by primers created by the primase. DNA polymerase I completes the Okazaki fragments, eliminating the RNA primer of the previous Okazaki fragment and replacing it with DNA. Finally, DNA ligase connects the fragments. Also involved in the process are single-strand binding proteins and topoisomerases that relieve the DNA's supercoiling. Initiation involves initiation proteins at *oriC,* and termination requires termination proteins bound to the termination sites and a topoisomerase.

**PROBLEM 4:** What can be concluded about the nucleic acids in the following table?

| Nucleic Acid Molecule | %A | %T | %G | %C | %U |
|---|---|---|---|---|---|
| a. | 28 | 28 | 22 | 22 | 0 |
| b. | 31 | 0 | 31 | 17 | 21 |
| c. | 15 | 15 | 35 | 35 | 0 |

***Answer:*** We must first look to see if U or T is present, for this will indicate whether the molecule is RNA or DNA, respectively. Molecule b is RNA; a and c are DNA. Now we look at base composition. In double-stranded molecules, A pairs evenly with T (or U) and G pairs with C. This relationship holds for molecules a and c, so they are double-stranded; molecule b is single-stranded. Finally, the melting temperature increases with the amount of G-C, so the melting temperature of c is greater than that of a.

## EXERCISES AND PROBLEMS*

### CHEMISTRY OF NUCLEIC ACIDS

1. If the tetranucleotide hypothesis were correct regarding the simplicity of DNA structure, under what circumstances could DNA be the genetic material?
2. Nucleic acids, proteins, carbohydrates, and fatty acids could have been mentioned as potential genetic material. What other molecular moieties (units) in the cell could possibly have functioned as the genetic material?
3. In what component parts do DNA and RNA differ?
4. Draw the structure of a short segment of DNA (three base pairs) at the molecular level. Indicate the polarity of the strands.
5. Roughly sketch the shape of B and Z DNA, remembering that B DNA is a right-handed helix and Z DNA is a left-handed helix.
6. Deduce whether each of the nucleic acid molecules in the following table is DNA or RNA and single-stranded or double-stranded.

| Nucleic Acid Molecule | %A | %G | %T | %C | %U |
|---|---|---|---|---|---|
| a. | 33 | 17 | 33 | 17 | 0 |
| b. | 33 | 33 | 17 | 17 | 0 |
| c. | 26 | 24 | 0 | 24 | 26 |
| d. | 21 | 40 | 21 | 18 | 0 |
| e. | 15 | 40 | 0 | 30 | 15 |
| f. | 30 | 20 | 15 | 20 | 15 |

7. A double-stranded DNA molecule is 28% guanosine (G).
   - **a.** What is the complete base composition of this molecule?
   - **b.** Answer the same question, but assume the molecule is double-stranded RNA.
8. The following are melting temperatures for five DNA molecules: 73° C, 69° C, 84° C, 78° C, 82° C. Arrange these DNAs in increasing order of percentage of G-C pairs.
9. We normally think that single-stranded nucleic acids should not melt, but many, in fact, do have a $T_m$. How can you explain this apparent mystery?
10. In a single-stranded DNA molecule, the amount of G is twice the amount of A, the amount of T is three times the amount of C, and the ratio of pyrimidines to purines is 1.5:1. What is the base composition of the DNA?
11. A double-stranded DNA measures 6.5 m in length. Approximately how many base pairs does it contain?

### DNA REPLICATION—THE PROCESS

12. Diagram the results that Meselson and Stahl would have obtained (a) if DNA replication were conservative and (b) if it were dispersive.
13. What type of photo would J. Cairns have obtained if DNA replication were conservative? Dispersive?

---

*Answers to selected exercises and problems are on page A-10.

### DNA REPLICATION—THE ENZYMOLOGY

**14.** Following is a section of a single strand of DNA. Supply a strand, by the rules of complementarity, that would turn this into a double helix. What RNA bases would primase use if this segment initiated an Okazaki fragment? In which direction would replication proceed?

5′-ATTCTTGGCATTCGC-3′

**15.** What is a primosome in *E. coli*? a replisome? What enzymes make up each? What is the relationship between these structures?

**16.** What are the differences between continuous and discontinuous DNA replication? Why do both exist?

**17.** Describe the synthesis of an Okazaki fragment.

**18.** Describe the enzymology of the origin, continuation, and termination of DNA replication in *E. coli.*

**19.** Can you think of any other mechanisms besides topoisomerase activity that could release supercoiling in replicating DNA?

**20.** Draw a diagram showing how topoisomerase II (gyrase) might work.

**21.** Retroviruses are single-stranded RNA viruses that insert their genomes into the host DNA during their life cycle. But only double-stranded DNA can be inserted into double-stranded DNA.

  **a.** Propose a mechanism that retroviruses could use to insert their genomes.
  **b.** What novel enzymes might such viruses require?

**22.** Propose a mechanism by which a single strand of DNA can make multiple copies of itself.

**23.** Progeria is a human disorder that causes affected individuals to age prematurely; a nine-year-old often resembles a sixty- to seventy-year-old individual in appearance and physiology. Suppose you extract DNA from a progeric patient and find mostly small DNA fragments rather than the expected long DNA molecules. What enzyme(s) might be defective in patients with progeria?

### REPLICATION STRUCTURES

**24.** Under what circumstances would you expect to see a DNA theta structure? D-loop? rolling-circle? bubbles? What function does each structure serve?

### EUKARYOTIC DNA REPLICATION

**25.** In developing sea urchins, just after fertilization, the cells divide every thirty to forty minutes. In the adult, the cells divide once every ten to fifteen hours. The amount of DNA per cell is the same in each case, but the DNA obviously replicates much faster in developing cells. Propose an explanation to account for the difference in replication time.

## CRITICAL THINKING QUESTIONS

**1.** Mutants are used to study various aspects of the phenotype and genotype. How can we study genes that are critically important in the functioning of an organism? For example, how do we study mutations in the gene for DNA polymerase III in *E. coli,* when changes in this gene are usually lethal? Remember, to study the genes in bacteria, we need the bacteria to grow and form colonies in order to be scored for their phenotypes.

**2.** DNA and RNA differ in two major ways: DNA has deoxyribose sugar, whereas RNA has ribose, and DNA has thymine, whereas RNA has uracil. Why might those differences exist other than accidents of evolution?

*Suggested Readings for chapter 9 are on page B-5.*

# 10

# GENE EXPRESSION

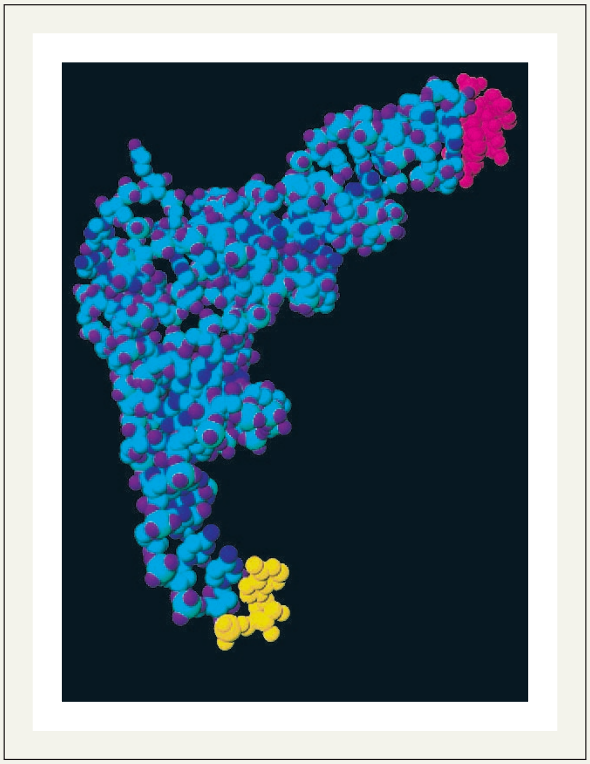
## *Transcription*

### STUDY OBJECTIVES

1. To examine the types of RNA and their roles in gene expression 245, 256
2. To look at the process of transcription, including start and stop signals, in both prokaryotes and eukaryotes 246
3. To investigate posttranscriptional changes in eukaryotic messenger RNAs, including an analysis of intron removal 260

### STUDY OUTLINE

**Types of RNA** 245
**Prokaryotic DNA Transcription** 246
  DNA-RNA Complementarity 246
  Prokaryotic RNA Polymerase 247
  Prokaryotic Initiation and Termination Signals for Transcription 248
**Ribosomes and Ribosomal RNA** 256
**Transfer RNA** 256
  Similarities of All Transfer RNAs 257
  Transfer RNA Loops 258
**Eukaryotic DNA Transcription** 260
  The Nucleolus in Eukaryotes 260
  Differences Between Eukaryotic and Prokaryotic Transcription 261
  Promoters 262
  Caps and Tails 265
  Introns 265
  RNA Editing 275
**Updated Information About the Flow of Genetic Information** 275
  Reverse Transcription 276
  RNA Self-Replication 276
  DNA Involvement in Translation 276
*Summary* 277
*Solved Problems* 277
*Exercises and Problems* 278
*Critical Thinking Questions* 279
Box 10.1 Observing Transcription in Real Time 250
Box 10.2 Polymerase Collisions: What Can a Cell Do? 252
Box 10.3 Are Viroids Escaped Introns? 272



A computer model of the serine transfer RNA. The amino acid binding site is *yellow;* the anticodon is *red.* (© Ken Eward/SPL/Photo Researchers.)

In this chapter, we continue our study of genetics at the molecular level. After discussing the structure of DNA and the way in which it replicates in the last chapter, we turn our attention here and in the next chapter to the way in which the genetic material—primarily DNA—expresses itself. In this chapter, we concentrate on the conversion of DNA information into RNA information, the first step in gene expression. In the next chapter, we look at the conversion of RNA information into proteins. Later chapters discuss the control of these processes. We begin with prokaryotes and later in the chapter discuss the conceptually similar but functionally more complex process in eukaryotes.

All living things synthesize proteins. In fact, the types of proteins that a cell synthesizes determine the kind of cell it is. Hence, the genetic material must determine the types and quantities of proteins a cell synthesizes. Proteins (polypeptides) are made up of strings of amino acids (three hundred to five hundred, on average) joined together by peptide bonds. (We cover protein structure and synthesis in chapter 11.) Each protein contains a unique combination of only twenty amino acids. The amino acid sequence is specified by the sequence of nucleotides in DNA or RNA. In all prokaryotes, eukaryotes, and DNA viruses, the gene is a sequence of nucleotides in DNA that codes for the sequence of RNA. That RNA then determines which amino acids are included in a polypeptide. RNA usually serves as an intermediary between DNA and proteins. (In RNA viruses, the RNA may serve as a template for the eventual synthesis of DNA, or the RNA may serve as genetic material without DNA ever being formed. We will consider these cases at the end of the chapter.)

In 1958, Francis Crick originally described the flow of genetic information as the **central dogma:** DNA transfers information to RNA, which then directly controls protein synthesis (fig. 10.1). DNA also controls its own replication. **Transcription** is the process of synthesizing RNA from a DNA template using the rules of complementarity—the DNA information is rewritten, but in the same nucleotide language. RNA controls the synthesis of proteins in a process called **translation** because the information in the language of nucleotides is *translated* into information in the language of amino acids.

In the previous chapter, we introduced the idea of proteins that recognize specific DNA sequences and bind to those sequences. Specifically, we introduced the initiator proteins that bind to *oriC* and the proteins that bind to the terminator sequences. DNA polymerases and some of the other proteins involved in DNA replication bind to DNA, but they do not necessarily bind to any specific sequences. Proteins that recognize specific DNA sequences are critically important to the transcriptional process. In the next chapter, we spend more time on proteins, discussing their structures and how they are synthesized. It is sufficient to say here that specific proteins recognize specific DNA sequences. They do so by interdigitating the amino acid side chains of the proteins into the grooves of the DNA, thereby recognizing specific sequences by hydrogen bonding and other electrostatic interactions between the side chains of the amino acids of the proteins and the bases of the DNA (fig. 10.2). Proteins can have parts that recognize DNA sequences and parts that recognize other proteins or that perform other enzymatic activities such as hydrolyzing ATP.

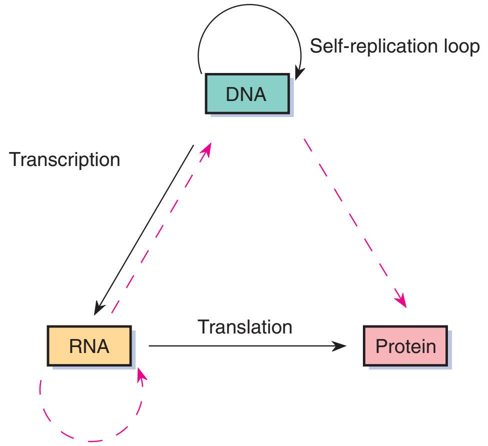


**Figure 10.1** Crick's original central dogma depicted the flow of genetic information. Dashed *red* lines indicate the possible information transfers unconfirmed in 1958, when Crick proposed the central dogma.

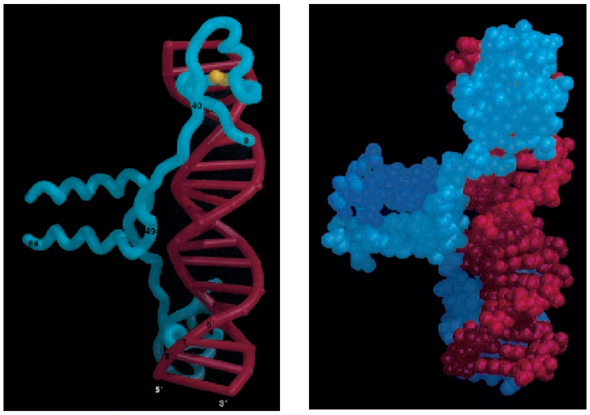
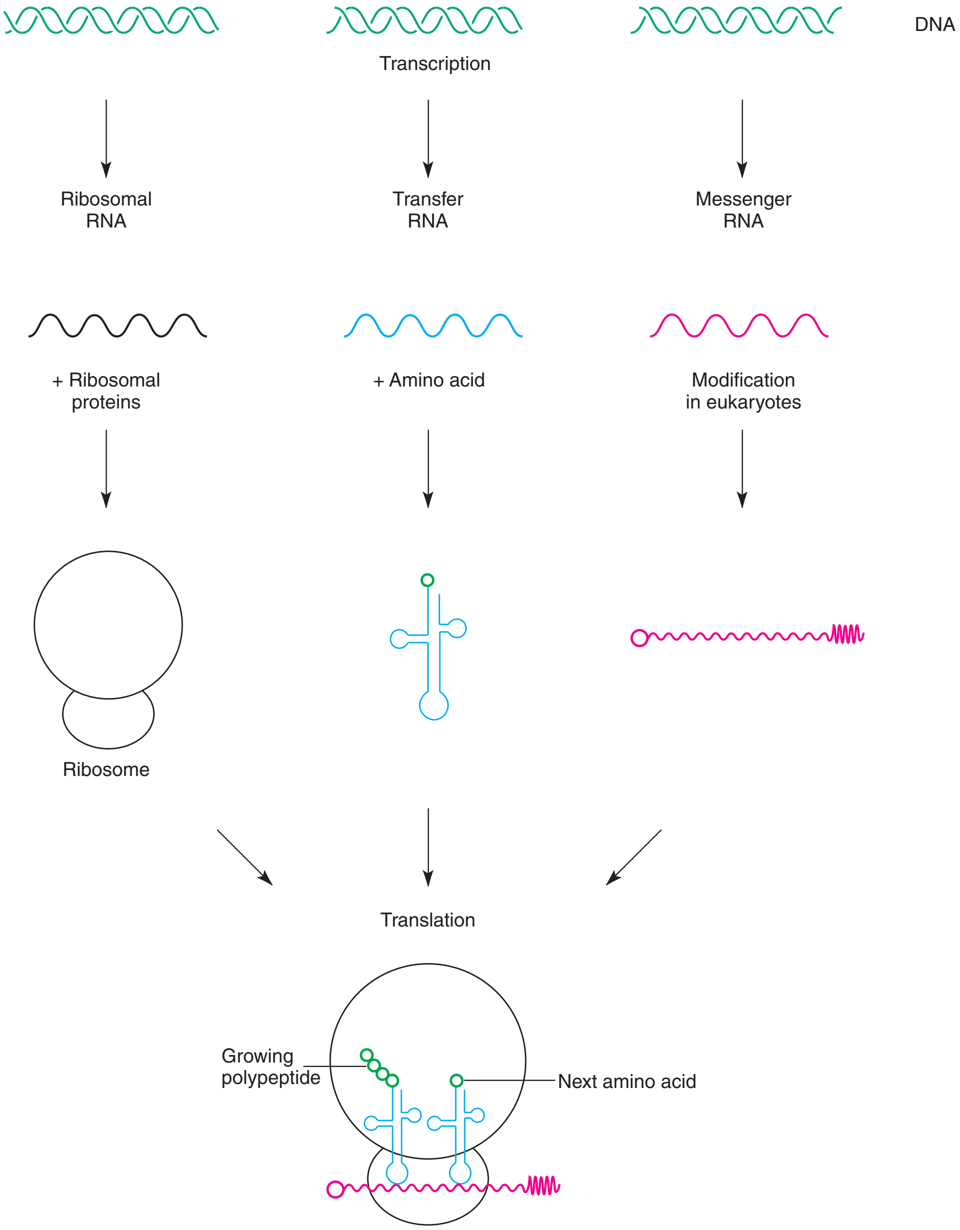

(a) (b)

**Figure 10.2** Computer model of the interaction of a yeast transcriptional factor, GAL4 (*blue*), and a seventeen-base-pair region of DNA (*red*). Zinc ions are in *yellow.* The protein is a dimer; only the DNA recognition region and associated part are shown. Part (*b*) is a space-filling model of part (*a*). (Reprinted with permission from *Nature,* 2 April 1992, Vol. 356, p. 411, fig. 3b,c. Copyright 1992 Macmillan Magazines Limited.)

## TYPES OF RNA

In the protein synthesis process, three different kinds of RNA serve in three different roles. The first type is **messenger RNA (mRNA),** which carries the DNA sequence information to particles in the cytoplasm known as **ribosomes,** where the messenger RNA is translated. The second type is **transfer RNA (tRNA),** which brings the amino acids to the ribosomes, where protein synthesis takes place. The third type of RNA is a structural and functional part of the ribosome called **ribosomal RNA (rRNA).** The general relationship of the roles of these three types of RNA is diagrammed in figure 10.3. In addition, small RNAs play other roles in cellular metabolism, some of which are described later in the chapter.



DNA
Transcription
Ribosomal RNA
Transfer RNA
Messenger RNA
+ Ribosomal proteins
+ Amino acid
Modification in eukaryotes
Ribosome
Translation
Growing polypeptide
Next amino acid

**Figure 10.3** Relationship among the three types of RNA—ribosomal, transfer, and messenger—during protein synthesis. All three types are found together at the ribosome during protein synthesis.

We know that DNA does not take part directly in protein synthesis because, in eukaryotes, translation occurs in the cytoplasm, whereas DNA remains in the nucleus. We suspected for a long time that the genetic intermediate in prokaryotes and eukaryotes was RNA because the cytoplasmic RNA concentration increases with increasing protein synthesis, and the cytoplasmic RNAs carry nucleotide sequences complementary to the cell's DNA. Proof of an RNA intermediate came when it was shown that messenger RNA directs protein synthesis.

## PROKARYOTIC DNA TRANSCRIPTION

### *DNA-RNA Complementarity*

What proof do we have that a messenger RNA exists? That is, what proof convinced geneticists that gene-sized RNAs (not transfer RNAs or ribosomal RNAs) were found in the cytoplasm that were complementary to the DNA in the nucleus? At least two lines of evidence exist. First, it was shown that the RNAs produced by various organisms have base ratios very similar to the base ratios in the same organisms' DNA (table 10.1). The second line of evidence comes from experiments by B. Hall, S. Spiegelman, and others using **DNA-RNA hybridization.** This technique denatures DNA by heating, which causes the two strands of the double helix to separate. When the solution cools, a certain proportion of the DNA strands rejoin and rewind—that is, complementary strands "find" each other and re-form double helices. When RNA is added to the denatured DNA solution and the solution is cooled slowly, some of the RNA forms double helices with the DNA if the RNA fragments are complementary to a section of the DNA (fig. 10.4). The existence of extensive complementarity between DNA and RNA is a persuasive indication that DNA acts as a template for complementary RNA.

In another experiment, DNA-RNA hybridization showed that bacteriophage infection led to the production of phage-specific messenger RNA. Gene-sized pieces of RNA extracted from *Escherichia coli* before and after bacteriophage T2 infection were tested to see if they hybridized with the DNA of the T2 phage or with the DNA of the *E. coli* cell. The RNA in the *E. coli* cell was found to hybridize with the *E. coli* DNA before infection but with the T2 DNA after infection. Thus it is apparent that when the phage attacks the *E. coli* cell, it starts to manufacture RNA complementary to its own DNA and stops the *E. coli* DNA from serving as a template.

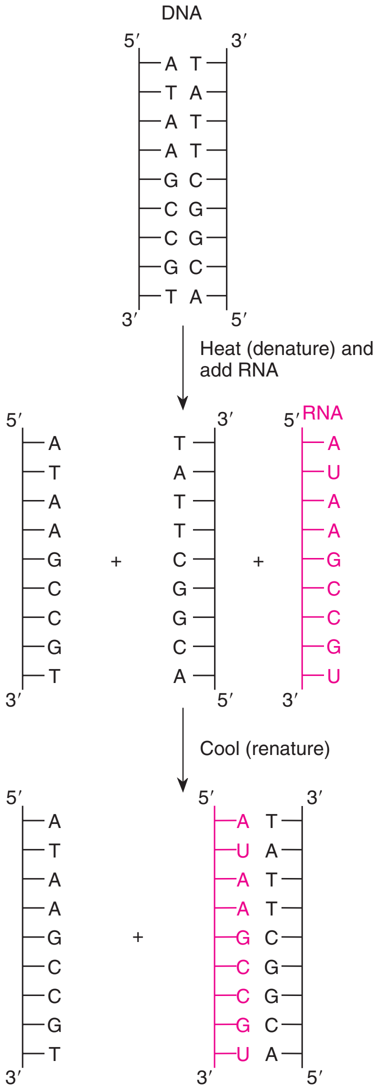


**Figure 10.4** DNA-RNA hybridization occurs between DNA and complementary RNA.

**Table 10.1** Correspondence of Base Ratios Between DNA and RNA of the Same Species

| | RNA % G + C | DNA % G + C |
|---|---|---|
| *E. coli* | 52 | 51 |
| T2 phage | 35 | 35 |
| Calf thymus gland | 40 | 43 |

Having reached the conclusion that RNA is transcribed (synthesized) from a DNA template and then directs protein synthesis, we look at two questions. First, is this RNA single- or double-stranded? Second, is it synthesized (transcribed) from one or both strands of the parental DNA?

For the most part, cellular RNA does not exist as a double helix. It can form double helical sections when complementary parts come into apposition (e.g., see fig. 10.16), but its general form is not a double helix. The simplest, and most convincing, evidence for this is that complementary RNA bases do not occur in corresponding proportions (Chargaff's ratios). That is, in RNA, uracil does not usually occur in the same quantity as adenine, nor does cytosine occur in the same quantity as guanine (table 10.2).

The answer to the second question is that RNA is not usually copied from both strands of any given segment of the DNA double helix, although rare exceptions do occur. Consider a sequence of nucleotides on one strand of a DNA duplex that specifies a sequence of amino acids for a protein, with the complementary nucleotide sequence also specifying the amino acid sequence for another functional protein. Since most enzymes are three hundred to five hundred amino acids long, the virtual impossibility of this task is obvious. It was, therefore, assumed *a priori* that, for any particular gene—that is, in any particular segment of DNA—the sequence on only one strand is transcribed and its complementary sequence is not. There is now considerable evidence to support this assumption.

The most impressive evidence that only one DNA strand transcribes RNA comes from work done with bacteriophage SP8, which attacks *Bacillus subtilis.* This phage has an interesting property—a great disparity in the purine-pyrimidine ratio of the two strands of its DNA. The disparity is significant enough that the two strands can be separated by density using density-gradient centrifugation. After denaturation and separation of the two strands, DNA-RNA hybridization can be carried out separately on each of the two strands with the RNA produced after the virus infects the bacterium. J. Marmur and his colleagues found that hybridization occurred only between the RNA and the heavier of the two DNA strands. Thus, only the heavy strand acted as a template for the production of RNA during the infection process.

The idea that only one strand of DNA serves as a transcription template for RNA has also been verified for several other small phages. However, when we get to larger viruses and cells, we find that either of the strands may be transcribed, but only one strand is used as a template in any one region. This was clearly shown in phage T4 of *E. coli,* where certain RNAs hybridize with one DNA strand, and other RNAs hybridize with the other. Let us now look at the transcription process in prokaryotes, then proceed to examine the three types of RNA in detail, and finally look at transcription in eukaryotes.

**Table 10.2** Base Composition in RNA (percentage)

| | Adenine | Uracil | Guanine | Cytosine |
|---|---|---|---|---|
| *E. coli* | 24 | 22 | 32 | 22 |
| *Euglena* | 26 | 19 | 31 | 24 |
| Poliovirus | 30 | 25 | 25 | 20 |

## Prokaryotic RNA Polymerase

In prokaryotes, transcription of RNA is controlled by **RNA polymerase.** Using DNA as a template, this enzyme polymerizes ribonucleoside triphosphates (RNA nucleotides). The complete RNA polymerase enzyme of *E. coli*—the holoenzyme—is composed of a core enzyme and a **sigma factor.** The core enzyme is composed of four subunits: $\alpha$ (two copies), $\beta$, and $\beta'$; this core is the component of the holoenzyme that actually carries out polymerization. The sigma factor is involved in recognizing transcription start signals on the DNA. Following the initiation of transcription, the sigma factor disassociates from the core enzyme.

Logically, transcription should not be a continuous process like DNA replication. If there were no control of protein synthesis, all the cells of a higher organism would be identical, and a bacterial cell would be producing all of its proteins all of the time. Since some enzymes depend on substrates not present all of the time, and since some reactions in a cell occur less frequently than others, the cell—be it a bacterium or a human liver cell—needs to regulate its protein synthesis. One of the most efficient ways for a cell to exert the necessary control over protein synthesis is to perform transcription selectively. Transcription of nongenic regions or of genes coding for unneeded enzymes is wasteful. Therefore, RNA polymerase should be selective. It should use as transcription templates only those DNA segments (genes or small groups of genes) whose products the cell needs at that particular time.

The mechanisms of transcriptional control need to be examined in two ways. First, we need to understand how the beginnings and ends of transcribable sections (a single gene or a series of adjacent genes) are demarcated. Second, we need to understand how the cell can selectively repress or enhance transcription of certain of these transcribable sections. The latter issues—the keys to bacterial efficiency and eukaryotic growth and development—are covered in chapters 14 and 16, respectively.

RNA polymerase must be able to recognize both the beginnings and the ends of genes (or gene groups) on the DNA double helix in order to initiate and terminate transcription. It must also be able to recognize the correct DNA strand to avoid transcribing the DNA strand that is not informational. RNA polymerase accomplishes those tasks by recognizing certain start and stop signals in DNA, called initiation and termination sequences, respectively.

## *Prokaryotic Initiation and Termination Signals for Transcription*

The DNA region that RNA polymerase associates with immediately before beginning transcription is known as the **promoter.** The promoter is an important part of gene expression in both prokaryotes and eukaryotes. Promoters contain the information for transcription initiation and are the major sites in which gene expression is controlled.

Without the sigma factor, the core enzyme of RNA polymerase binds randomly along the DNA. Formation of the holoenzyme brings about high affinity of RNA polymerase for DNA sequences in the promoter region. Termination of transcription comes about when the polymerase enzyme recognizes a DNA region known as a **terminator sequence.** Let us elaborate on the various stages of transcription (in this section and in boxes 10.1 and 10.2).

### *Promoters*

The RNA polymerase molecule covers a region of about sixty base pairs of DNA. This was determined by causing the polymerase to bind to DNA and then digesting the mixture with nucleases, in a technique known as **footprinting** (fig. 10.5). The polymerase "protects" or prevents degradation of the region it covers. The undigested DNA is then isolated and its size determined. Geneticists have gained much new information about the nature of recognition regions within promoters through recombinant DNA technology and nucleotide sequencing techniques (see chapter 13). Sequencing of numerous promoters has shown that they contain common sequences. If the promoter nucleotide sequences align with each other, and each has exactly the same series of nucleotides in a given segment, we say that the sequence of that segment comprises an invariant or **conserved sequence.** If, however, there is some variation in the sequence, but certain nucleotides occur at a high frequency (significantly greater than by chance), we refer to those nucleotides as making up a **consensus sequence.** Surrounding a point in prokaryotic promoters about ten nucleotides before the first transcribed base is just such a consensus sequence—TATAAT. This sequence is known as a **Pribnow box** after one of its discoverers (fig. 10.6).

The nucleotides in the Pribnow box are mostly adenines and thymines, so the region is primarily held together by only two hydrogen bonds per base pair. Since local DNA denaturation occurs during transcription by RNA polymerase (the DNA is opened to allow transcription), fewer hydrogen bonds make this process easier energetically. When the polymerase is bound at the promoter region (fig. 10.6), it is in position to begin polymerization six to eight nucleotides down from the Pribnow box.

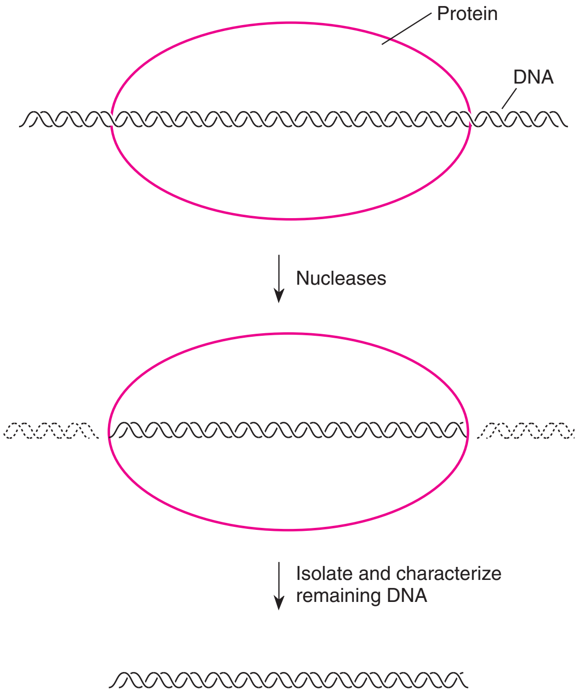


**Figure 10.5** Footprinting technique. DNA in contact with a protein (e.g., RNA polymerase) is protected from nuclease degradation. The protected DNA is then isolated and characterized.

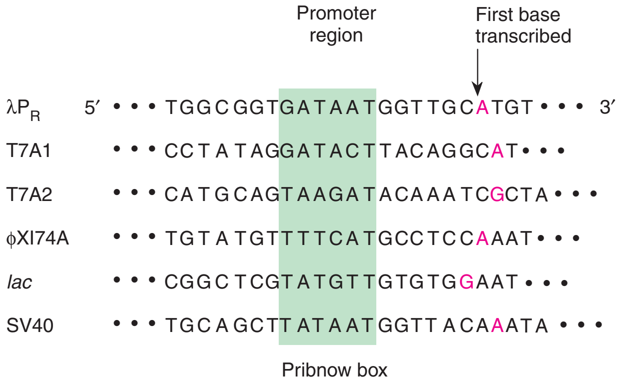

Promoter region
First base transcribed
λP_R 5′ • • • TGGCGGTGATAATGGTTGCATGT • • • 3′
T7A1 • • • CCTATAGGATACTTACAGGCAT • • •
T7A2 • • • CATGCAGTAAGATACAAATCGCTA • • •
φXI74A • • • TGTATGTTTTCATGCCTCCAAAT • • •
lac • • • CGGCTCGTATGTTGTGTGGAAT • • •
SV40 • • • TGCAGCTTATAATGGTTACAAATA • • •
Pribnow box

**Figure 10.6** Nucleotide sequences of the promoter region and the first base transcribed from several different genes. Lambda (λ), T7, and φ×174 are bacteriophages. *Lac* is an *E. coli* gene, and SV40 is an animal virus. Only the SV40 promoter has the actual consensus sequence of TATAAT. Even when other sequenced promoters not shown here are Included, no base is found 100% of the time (conserved).

The sequences shown in figure 10.6 are those of the **coding strand** of DNA. It is a general convention to show the coding strand because both that strand and messenger RNA have the same sequences, substituting U for T in RNA; they are both complementary to the same **template strand** (also referred to as the **anticoding strand** or **noncoding strand;** fig. 10.7). Another convention is to indicate the first base transcribed by number +1 and to use positive numbers to count farther down the DNA in the **downstream** direction of transcription. If transcription is proceeding to the right, the direction to the left is called **upstream,** with bases indicated by negative numbers (fig. 10.8). Under this convention, the Pribnow box is often referred to as the −10 sequence.

Figure 10.8 also indicates another region with similar sequences among many promoters centered near −35 and referred to as the −35 sequence. The consensus sequence at −35 is TTGTCA. Mutation studies have attempted to determine the relative roles of the −10 and −35 sequences in transcription. In other words, mutations of bases in the −10 and −35 regions were examined to determine how they affected transcription initiation. The conclusions from these studies are that both regions contribute to the efficiency of polymerase binding. In other words, the more each sequence differs from the consensus sequence, the less frequently that promoter initiates transcription. The sigma factor recognizes both the −35 and the −10 sequences. The sigma factor is also sensitive to the spacing between these sequences, preferring (being most efficient at) seventeen base pairs.

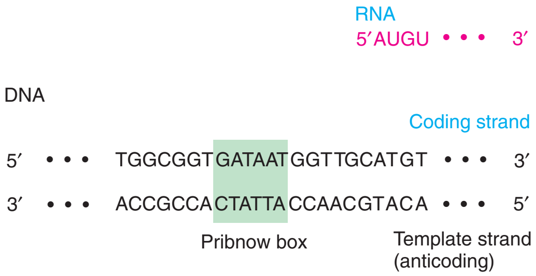

RNA
5′AUGU • • • 3′
DNA
Coding strand
5′ • • • TGGCGGTGATAATGGTTGCATGT • • • 3′
3′ • • • ACCGCCACTATTACCAACGTACA • • • 5′
Pribnow box
Template strand (anticoding)

**Figure 10.7** The template (anticoding) strand of DNA is complementary to both the coding strand and the transcribed RNA. The sequences are from the promoter of the $\lambda P_R$ region (see fig. 10.6).

Farther upstream from the −35 sequence is a recognition element in bacterial promoters that are very strongly expressed, such as the ribosomal RNA genes (fig. 10.8). This **upstream element,** or **UP element,** is about twenty base pairs long, is centered at −50, and is rich in A and T. By mutational studies, it has been shown that adding this element to promoters that don't normally have it greatly increases the rate of transcription. There are other recognition sites in prokaryotes, both upstream and downstream, at which various proteins attach that can enhance or inhibit transcription by direct contact with the polymerase (the α and σ subunits). We discuss these in chapter 14 under control of transcription. They are not part of what we think of as the core promoter, the DNA sequence needed for efficient binding of RNA polymerase.

Since the holoenzyme recognizes consensus sequences in a promoter, it is not surprising that some promoters are bound more efficiently than others or that different sigma factors exist within a cell. In *E. coli,* the major sigma factor is a protein of 70,000 daltons, referred to as $\sigma^{70}$. (One dalton is an atomic mass of 1.0000, approximately equal to the mass of a hydrogen atom.) The existence of about five less common sigma factors provides the cell with a mechanism for transcribing different genes under different circumstances. For example, in an *E. coli* cell subjected to elevated temperatures, a group of new proteins, referred to as **heat shock proteins,** appear, acting to protect the cell to some extent against the elevated temperatures. These proteins all appear at once because they have promoters that a different sigma factor recognizes, one with a molecular weight of 32,000 daltons ($\sigma^{32}$); this new sigma factor is produced by the cell after

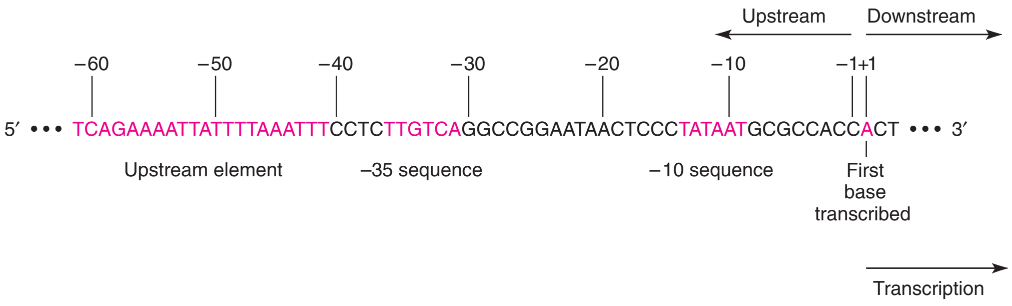


**Figure 10.8** Promoter of the *Escherichia coli* ribosomal RNA gene, *rrnB.* Note the −10 and −35 sequences and the upstream element. The first base transcribed (the transcriptional start site) is noted (+1), as well as the upstream, downstream, and transcription directions. (Data from W. Ross, et al., 1993. *Science* 262:1407.)

## BOX 10.1

## Experimental Methods

### *Observing Transcription in Real Time*

The overwhelming evidence that molecular events, such as transcription, take place comes from genetic and biochemical analyses and occasionally an electron micrograph of one type or another (fig. 1). Thus, it is refreshing and illuminating to be able to observe some of the processes we know are taking place in real time; that is, to sit at a microscope and actually see these events happen. Such a study on transcription was published in 1991 in *Nature* by four scientists at Washington University in St. Louis.

Although new methods of microscopy are being developed, normally we cannot see these molecular events taking place; the components are too small. Making them visible in electron microscopes usually requires fixation that destroys the ability of the components to actually continue their tasks. The Washington University group overcame this by attaching a gold particle to DNA, thus rendering the motion of that DNA visible under the light microscope (fig. 2). The scientists immobilized the RNA polymerase to a glass coverslip; thus, as transcription took place, the DNA moved and the length of the tether of the gold particle increased. At first they stopped the process by limiting the concentration of nucleoside triphosphates (NTPs). They

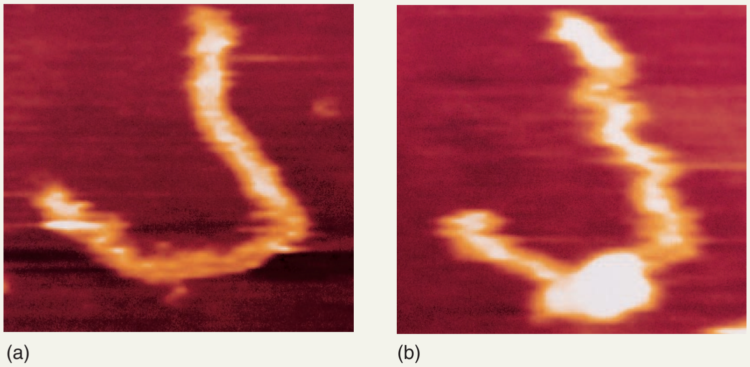


**Figure 1** Visualizing transcription. Image of DNA before (*a*) and after (*b*) *E. coli* RNA polymerase (bright oblong object in *b*) binds to a promoter. Pictures are by scanning force microscopy, a new laser technique that images molecules in water. Image sizes are 300 by 300 nm. *Dark brown* represents substrate level; the highest point is *white* at about 10 nm high. Intermediate colors represent intermediate heights. (Courtesy of Martin Guthold and Carlos Bustamante, Institute of Molecular Biology and HHMI, University of Oregon.)

heat shock. We discuss heat shock proteins and other systems of transcriptional control in chapters 14 and 16.

From mutational studies of promoters and the proteins in the RNA polymerase holoenzyme, we now have a picture of a holoenzyme that sets down on a DNA promoter because the sigma factor recognizes the $-10$ and $-35$ elements, the $\alpha$ proteins recognize the UP element, and the $\alpha$ and $\sigma$ subunits recognize proteins bound to various other upstream elements, when present (fig. 10.9*a*). This initiation complex is initially referred to as a closed complex because the DNA has not melted, which is the next step in transcription initiation (fig. 10.9*b*). After the transcription of 5–10 bases, the sigma factor is released (fig. 10.9*c* and *d*).

About seventeen base pairs of DNA are opened, and as transcription proceeds, about twelve bases of RNA form a DNA-RNA duplex at the point of transcription. Some of this information comes from studies with potassium permanganate ($KMnO_4$), which modifies DNA bases that are single-stranded but not double-stranded. Thus, the lengths of melted DNA can be determined experimentally. Also used is the technique of **photocrosslinking,** in which two moieties such as DNA and one or two proteins are caused to be permanently crosslinked, verifying their close contact. This is done by attaching a chemical crosslinking element to one of the moieties and then causing crosslinking to occur by shining light, usually ultraviolet, on the mixture.

Transcription, like DNA replication, always proceeds in the $5' \rightarrow 3'$ direction. That is, a single base is added *de novo* and then new RNA nucleotides are added to the $3'$-OH free end, as in DNA replication. However, unlike could then observe the motion of the gold ball when no transcription was taking place. The scientists predicted that an immobilized gold ball would not move, and a tethered gold ball would show a limited amount of Brownian motion. That is, it would show a limited amount of blur in light microscope video images averaged over time. However, as soon as NTPs were added, any tethered gold ball would show an increased blur as it moved out of the field of vision and eventually would be released when transcription was completed. That is exactly what they saw (fig. 3). Thus, they succeeded in watching transcription take place in real time.

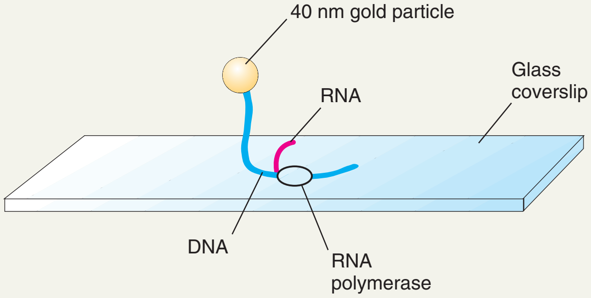


**Figure 2** An experimental design in visualizing transcription in real time under the light microscope. Here, an RNA polymerase is immobilized on a coverslip, waiting for nucleoside triphosphates (NTPs) to be added. The gold particle is tethered by the DNA, allowing us to keep track visually of the end of the DNA.

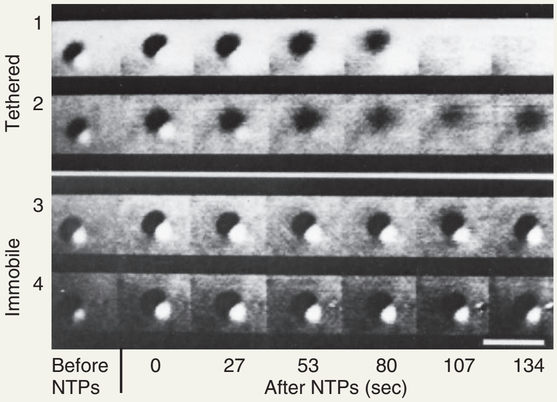


**Figure 3** Enhanced light microscope images of the gold particles. In *3* and *4,* presumably immobilized particles show no change in focus over time. In *1* and *2,* Brownian motion—and hence blur—increases through time consistent with the lengthening of the tether (transcription). The particle in *1* was released 87 seconds (and in *2,* 135 seconds) after NTPs were added. Scale bar is 1 μm. (Transcription by single molecules of RNA polymerase observed by light microscopy. Robert Landick, Department of Biology, Washington University, St. Louis, MO.)

DNA polymerase, prokaryotic RNA polymerase does not seem to proofread as it proceeds. That is, RNA polymerase evidently does not verify the complementarity of the new bases added to the growing RNA strand. This deficiency is not serious; since many messenger RNAs are short-lived and many copies are made from actively transcribed genes, an occasional mistake will probably not produce permanent or overwhelming damage. If a particular RNA is not functional, a new one will be made soon. Evolutionarily speaking, it seems that it is more important to make RNA quickly than to proofread each RNA made.

### Terminators

Transcription continues as RNA polymerase adds nucleotides to the growing RNA strand according to the rules of complementarity (C, G, A, and U of RNA pairing with G, C, T, and A of DNA, respectively). The polymerase moves down the DNA until the RNA polymerase reaches a stop signal, or terminator sequence. Two types of terminators, rho-dependent and rho-independent, differ in their dependency on the **rho protein** (Greek letter ρ). The functional form of rho is a hexamer, six identical copies of the protein. **Rho-independent terminators** cause termination of transcription even if rho is not present. **Rho-dependent terminators** require the rho protein; without it, RNA polymerase continues to transcribe past the terminator in a process known as **read-through.** Both types of terminators sequenced so far have one thing in common: They include a sequence and its inverted form separated by another short sequence, all together forming an **inverted-repeat sequence.** The

## BOX 10.2

## Experimental Methods

### *Polymerase Collisions: What Can a Cell Do?*

Both RNA polymerase and DNA polymerase move along the DNA of a cell during a cell cycle. The DNA polymerase moves along at about ten times the speed of the transcribing enzyme. Since many genes are usually active in a cell, the interaction (collision) of the two enzymes is inevitable. What happens when this collision takes place? What does the cell do? Although we cannot directly observe these interactions, various bits of data suggest that a head-on collision could be fatal to the cell, and certain patterns of gene placement minimize the chance of a head-on collision.

B. Brewer first analyzed the problem of the coexistence of these two enzymes in a paper published in 1988 in the journal *Cell.* In evolutionary terms, the cell could obviate the problems of a head-on collision by either avoiding them or resolving them. Resolution would entail some sort of right-of-way settlement when the two enzymes met; for example, the RNA polymerase could drop off the DNA when a confrontation takes place. The cell might avoid confrontations if the genes are oriented so that transcription occurs for the most part in the same direction as DNA replication. That is, DNA replication begins at *oriC,* with Y-junctions proceeding to the left and the right until they meet 180 degrees later. Thus, to avoid head-on collisions, genes on the left and right arcs of the bacterial chromosome could be transcribed away from the origin of replication (fig. 1).

Brewer analyzed the orientation of genes on the *E. coli* chromosome; more recently, D. Zeigler and D. Dean did the same for the chromosome of *Bacillus subtilis.* In *B. subtilis,* 95% (91 of 96) of the genes analyzed were in the proper orientation to avoid a head-on collision of polymerases. Among the exceptions were sporulation genes, genes that would not be transcribed during DNA synthesis and whose orientation is thus not relevant to DNA polymerase activity. In *E. coli,* Brewer found that, overall, 74% (375 of 501) of the genes she looked at were oriented to avoid head-on collisions. Brewer's data were more impressive when she broke them down according to transcription function and activity.

For genes that transcribe very actively most of the time, the orientation is about 90% in the "safe" direction. For regulatory genes that are transcribed only very rarely, the orientation is random (50% safe). For other genes, the orientation was 72% in the safe direction. Thus, an organization clearly exists within the bacterial chromosome that helps to avoid head-on collisions of the two polymerases.

Brewer also provided evidence that a head-on collision between polymerases could be fatal to the cell. Studies selected inversions of the *E. coli* chromosome to see the effects of collision. (Inversions are regions that have been cut out and put back in the opposite orientation.) It was impossible to isolate inversion mutations that changed the orientation of genes in respect to *oriC.* Thus, it appears that a cell may not be able to resolve a head-on collision of polymerases and that evolution has solved the problem by having gene transcription generally oriented in the same direction as DNA replication.

More amazingly, Alberts and his colleagues recently studied what happens when a replication fork catches up to a stalled RNA polymerase. Not only does the replication fork pass the transcription apparatus, but the RNA polymerase can resume transcription after the replication fork passes without loss of the transcript. Although there are contrary observations in other systems, it appears that gene orientation and the behavior of polymerases allow cells to survive with both replication and transcription occurring on the same DNA.

---

terminator in figure 10.10 has the sequence AAAGGCTCC, 5′ → 3′, from both the left on the coding strand and from the right on the template strand. A four-base-pair sequence separates the inverted repeats. Inverted repeats can form a **stem-loop structure** by pairing complementary bases within the transcribed messenger RNA.

Both rho-dependent and rho-independent terminators have the stem-loop structure in RNA just before the last base transcribed. Rho-independent terminators, as figure 10.10 shows, also have a sequence of thymine-containing nucleotides after the inverted repeat, whereas rho-dependent terminators do not. Although the exact sequence of events at the terminator is not fully known, it appears that the RNA stem-loop structure forms and causes the RNA polymerase to pause just after completing it. This pause may then allow termination under two different circumstances.

In rho-independent terminators, the pause may occur just after the sequence of uracils is transcribed (fig. 10.11). Uracil-adenine base pairs have two hydrogen bonds and are thus less stable thermodynamically than guanine-cytosine base pairs. Perhaps during the pause, the uracil-adenine base pairs spontaneously denature, releasing the transcribed RNA and the RNA polymerase,

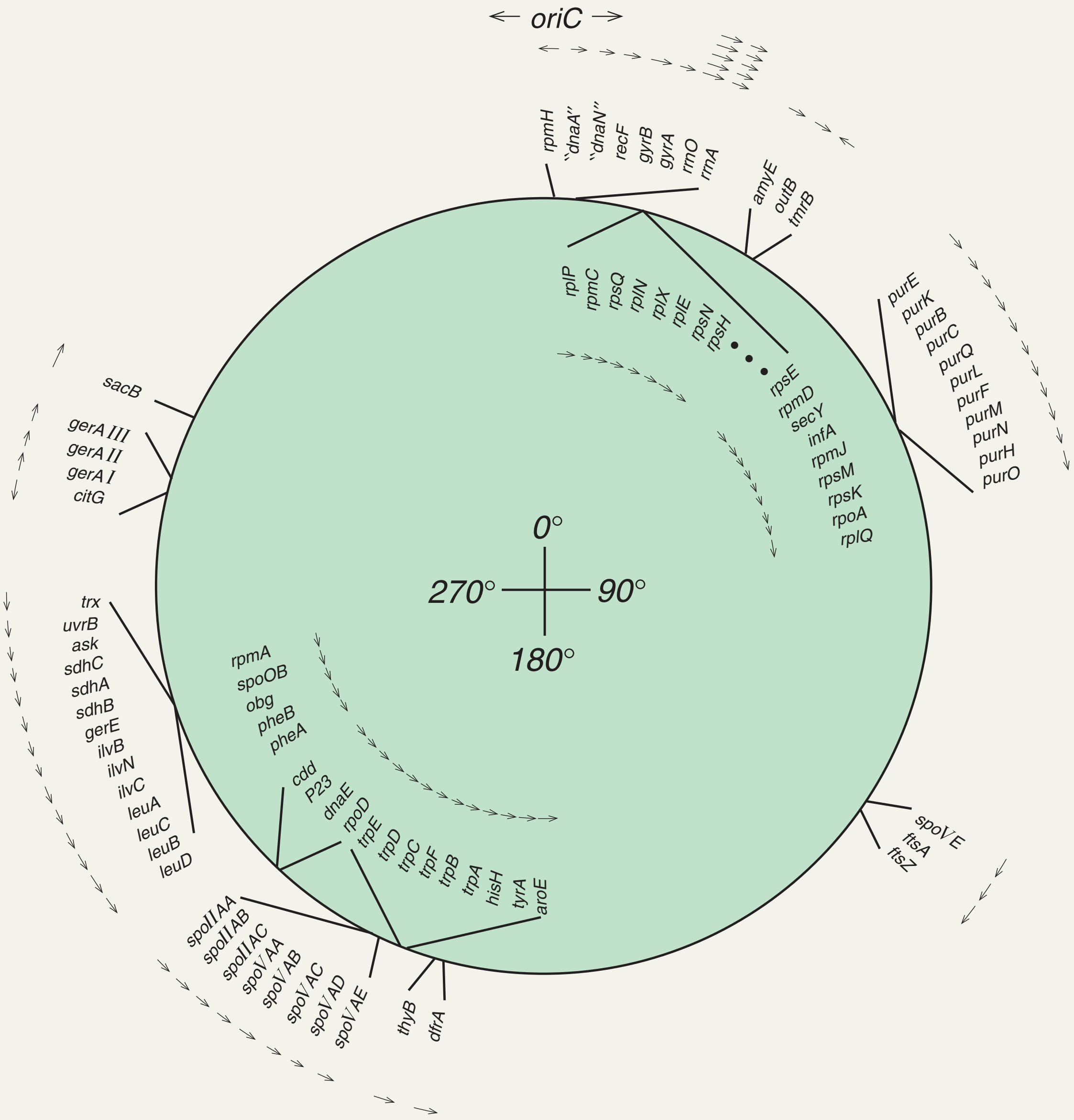

oriC
rpmH
"dnaA"
"dnaN"
recF
gyrB
gyrA
rrnO
rrnA
amyE
outB
tmrB
rplP
rpmC
rpsQ
rplN
rplX
rplE
rpsN
rpsH
rpsE
rpmD
secY
infA
rpmJ
rpsM
rpsK
rpoA
rplQ
purE
purK
purB
purC
purQ
purL
purF
purM
purN
purH
purO
sacB
gerA III
gerA II
gerA I
citG
0°
270°
90°
180°
trx
uvrB
ask
sdhC
sdhA
sdhB
gerE
ilvB
ilvN
ilvC
leuA
leuC
leuB
leuD
rpmA
spoOB
obg
pheB
pheA
cdd
P23
dnaE
rpoD
trpE
trpD
trpC
trpF
trpB
trpA
hisH
tyrA
aroE
spoVE
ftsA
ftsZ
spoIIAA
spoIIAB
spoIIAC
spoVAA
spoVAB
spoVAC
spoVAD
spoVAE
thyB
dfrA

**Figure 1** Location and orientation of gene transcription on the chromosome of *Bacillus subtilis* (*arrows*). DNA replication begins at *oriC* and terminates approximately 180 degrees from the origin of replication. Note that the overwhelming number of arrows point away from the origin of replication toward the termination point. (From D. R. Zeigler and D. H. Dean, "Orientation of genes in the *Bacillus subtilis* chromosome," *Genetics,* 125:703–8. Copyright © 1990 Genetics Society of America.)

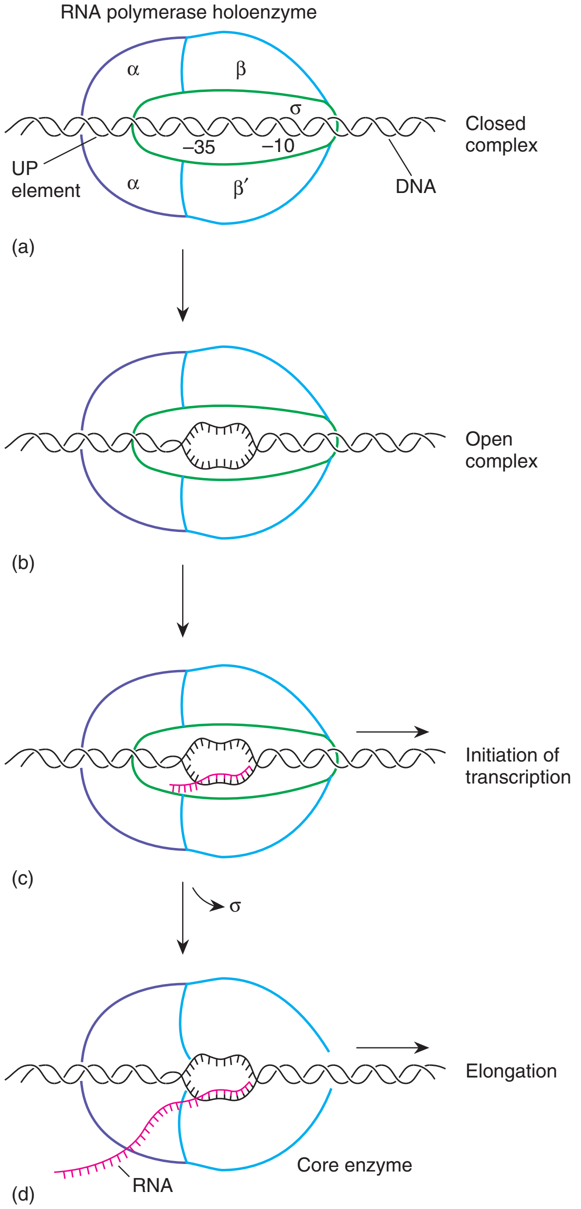

RNA polymerase holoenzyme
α
β
σ
−35
−10
UP element
α
β′
DNA
Closed complex
(a)
Open complex
(b)
Initiation of transcription
(c)
σ
Elongation
RNA
Core enzyme
(d)

**Figure 10.9** Transcription begins after RNA polymerase attaches to the promoter, with specificity imparted by the sigma factor. The DNA opens to form the open complex, transcription begins, the sigma factor leaves, and elongation commences.

terminating the process, and making the polymerase available for further transcription of other promoters.

Rho-dependent terminators do not have the uracil sequence after the stem-loop structure. Here, termination depends on the action of rho, which appears to bind to the newly forming RNA. In an ATP-dependent process, rho travels along the RNA at a speed comparable to the transcription process itself (fig. 10.11). Possibly, when RNA polymerase pauses at the stem-loop structure, rho catches up to the polymerase and unwinds the DNA-RNA hybrid, leting the DNA, RNA, and polymerase fall free. Rho can do this because it has DNA-RNA helicase (unwinding) properties.

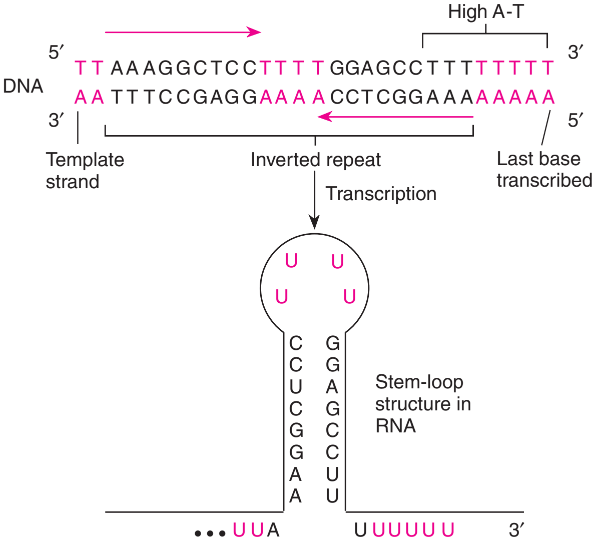

High A-T
5′
3′
DNA
TTAAAGGCTCCTTTTGGAGCCTTTTTTTT
AATTTCCGAGGAAAACCTCGGAAAAAAAA
3′
5′
Template strand
Inverted repeat
Last base transcribed
Transcription
U U
U U
C G
C G
U A
C G
G C
G C
A U
A U
Stem-loop structure in RNA
•••UUA
UUUUUU
3′

**Figure 10.10** An inverted-repeat base sequence characterizes terminator regions of DNA. Stem-loop structures can occur as the RNA forms because of complementary sequences. The 3′ poly-U tail indicates a rho-independent terminator.

The process of transcription termination is probably more complex than described. Significant interactions may take place with other proteins, and particular sequences surrounding the termination sequence may also be significant in the termination process. This is an area of active research.

Figure 10.12 shows an overview of transcription. The information of a gene, coded in the sequence of nucleotides in the DNA, has been transcribed into a complementary sequence of nucleotides in the RNA. This RNA transcript contains a complement of the template strand of the gene's DNA and thus acts as a messenger from the gene to the cell's protein-synthesizing complex. The transcript contains nucleotide sequences that will be translated into amino acids—coding segments—as well as noncoding segments before and after. The translatable segment, or gene, almost always begins with a three-base sequence, AUG, which is known as an initiator codon, and ends with one of the three-base sequences, UAA, UAG, or UGA, known as nonsense codons. (We discuss these signals in chapter 11.)

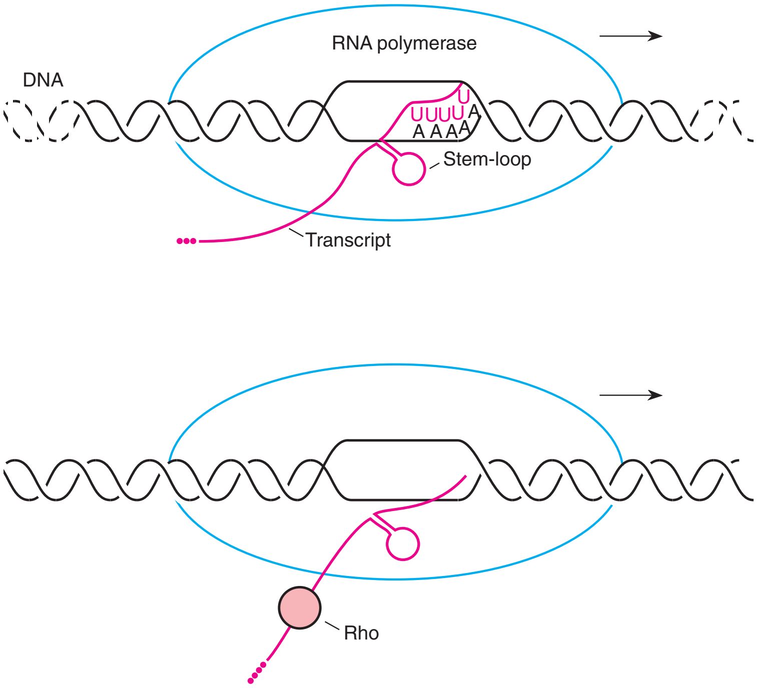


**Figure 10.11** Rho-independent (*top*) and rho-dependent (*bottom*) termination of transcription are preceded by a pause of the RNA polymerase at a terminator sequence. Presumably, the stem-loop structure in the nascent RNA causes the pause in both cases.

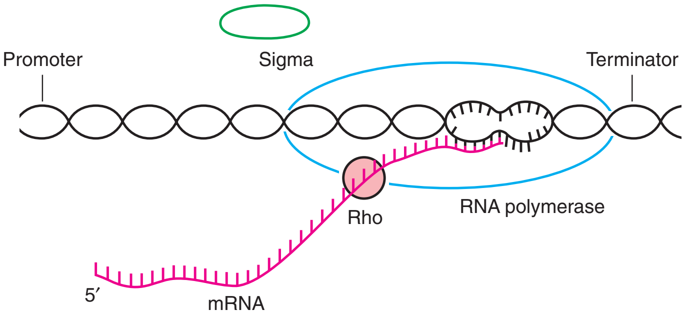


**Figure 10.12** Transcription overview and RNA polymerase molecules. RNA polymerase is transcribing near the terminator. The rho factor—actually made up of six subunits—is shown on the newly formed RNA. The sigma factor is shown nearby, detached from the core polymerase.

The portion of the RNA transcript that begins at the start of transcription and goes to the translation initiator codon (AUG) is referred to as a **leader,** or 5′ untranslated sequence. The length of RNA from the nonsense codon (UAA, UAG, or UGA) to the last nucleotide transcribed is the **trailer,** or 3′ untranslated sequence. These sequences play a role in recognizing messenger RNA and ensuring its structural stability at the ribosome during the process of translation; the leader region can also have regulatory functions (see chapter 14). Figure 10.13 diagrams a complete prokaryotic RNA transcript. In this simplified drawing, the transcript has only one gene (AUG → UAA). However, the average prokaryotic transcript contains the information for several genes. We will say more about the parts of a transcript later in this chapter and the next. Now we turn our attention to the types of transcripts: ribosomal, transfer, and messenger RNA.

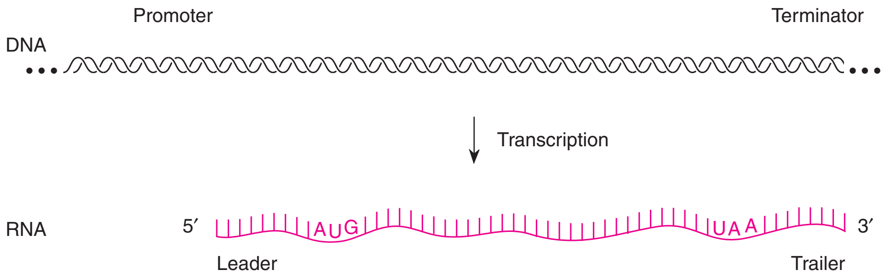


**Figure 10.13** Transcribed piece of prokaryotic RNA and its DNA template region. Note the promoter and terminator regions on the DNA and the leader and trailer regions on the RNA. The initiation (AUG) and nonsense (UAA) codons for protein synthesis are shown. These signals are read at the ribosome at the time of translation.

## RIBOSOMES AND RIBOSOMAL RNA

Ribosomes are organelles in the cell, composed of proteins and RNA (ribosomal RNA, or rRNA), where protein synthesis occurs. In a rapidly growing *E. coli* cell, ribosomes can make up as much as 25% of the mass of the cell. Ribosomes, as well as other small particles and molecules, are measured in units that describe their rate of sedimentation during density-gradient centrifugation in sucrose. This technique gives information on size and shape (due to the speed of sedimentation) while simultaneously isolating the molecules. Isolation by centrifugation in sucrose is a relatively gentle isolation technique; the molecules still retain their biological properties and can be used for further experimentation. In the 1920s, physical chemist T. Svedberg developed ultracentrifugation, giving his name to the unit of sedimentation: the **Svedberg unit,** S.

In sucrose density-gradient centrifugation, the gradient is formed by layering on decreasingly concentrated sucrose solutions. In a related technique, cesium chloride density-gradient centrifugation, mentioned in chapter 9, the gradient develops during centrifugation. The sucrose centrifugation is stopped after a fixed time, whereas in the cesium chloride technique, the system spins until it reaches equilibrium. The sucrose method tends to be more rapid. Samples can be isolated from a sucrose gradient by punching a hole in the bottom of the tube and collecting the drops in sequentially numbered containers. The first (lowest-numbered) containers will contain the heaviest molecules (with the highest S values).

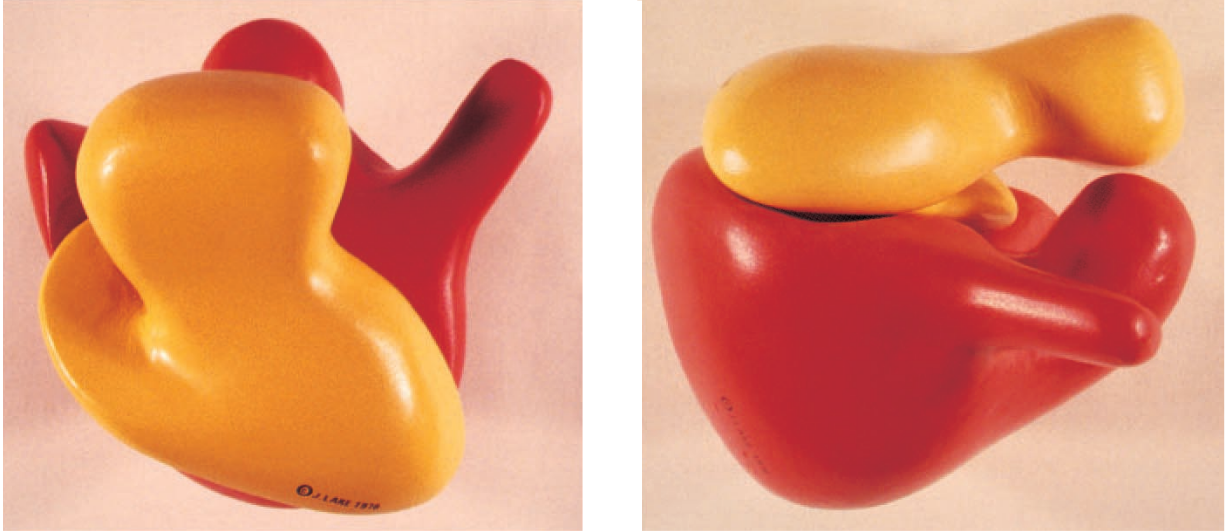
Ribosomes in all organisms are made of two subunits of unequal size. The sedimentation value is 50S (Svedberg units) for the large one in *E. coli* and 30S for the smaller one. Together they sediment at about 70S. Eukaryotic ribosomes vary from 55S to 66S in animals and 70S to 80S in fungi and higher plants. Most of our discussion will be confined to the well-studied ribosomes of *E. coli.*

Each ribosomal subunit comprises one or two pieces of ribosomal RNA and a fixed number of proteins. The 30S subunit of *E. coli* has twenty-one proteins and a 16S molecule of ribosomal RNA, and the 50S subunit has thirty-four proteins and two pieces of ribosomal RNA—one 23S and one 5S section (fig. 10.14). Advances in understanding ribosomal structure have come about after protein chemists isolated and purified all the proteins of the ribosome. This allowed researchers to experiment on the proper sequence needed to assemble the subunits and also allowed them to develop immunological techniques to show the positions of many proteins in the completed ribosomal subunits.

In *E. coli,* all three ribosomal RNA segments are transcribed as a single long piece of RNA that is then cleaved and modified to form the final three pieces of RNA (16S, 23S, and 5S). The region of DNA that contains the three ribosomal RNA molecules also contains genes for four transfer RNAs (fig. 10.15). There appear to be about five to ten copies of this region in each chromosome of *E. coli.* The occurrence of the three ribosomal RNA segments on the same piece of RNA ensures a final ratio of 1:1:1, the ratio needed for ribosomal construction.

## TRANSFER RNA

During protein synthesis (see fig. 10.3), a messenger RNA, carrying the information transcribed from the gene (DNA), is bound to the ribosome. Amino acids are brought to the ribosome attached to transfer RNAs. The code is read in sequences of three nucleotides, called **codons.** The nucleotides of the codon on messenger RNA are complementary to and pair with a sequence of three bases—the **anticodon**—on a transfer RNA. Each different transfer RNA carries a specific amino acid. Thus, the transfer RNA recognizes the specificity of the genetic code (fig. 10.16).



(a) **30S**
Twenty-one proteins
One 16S rRNA

(b) **50S**
Thirty-four proteins
One 23S rRNA
One 5S rRNA

**Figure 10.14** The *E. coli* ribosome. (*a*) and (*b*) show models of the 70S ribosome of *E. coli,* revealing the relationship of the small (*yellow*) and large (*red*) subunits at the time of translation. The 30S ribosomal subunit is composed of twenty-one proteins and one 16S piece of ribosomal RNA. The 50S subunit is composed of thirty-four proteins and two pieces of ribosomal RNA, 23S and 5S. ([*a* and *b*] James A. Lake, *Journal of Molecular Biology* 105 (1976):131–59. Reproduced by permission of Academic Press.)

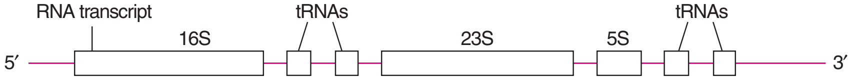


**Figure 10.15** The *E. coli* transcript that contains the three ribosomal RNA segments also contains four tRNAs and some spacer RNA (*red*) that separates the tRNA and rRNA genes.

The correct amino acid is attached to its transfer RNA by one of a group of enzymes called **aminoacyl-tRNA synthetases.** One specific aminoacyl synthetase exists for every amino acid, but the synthetase may recognize more than one transfer RNA because there are more transfer RNAs (and codons) than there are amino acids. (In chapter 11 we discuss the genetic code in more detail.) R. W. Holley, a Nobel laureate, and his colleagues were the first to discover the nucleotide sequence of a transfer RNA; in 1964, they published the structure of the alanine transfer RNA in yeast (fig. 10.17). The average transfer RNA is about eighty nucleotides long.

## *Similarities of All Transfer RNAs*

Transfer RNAs have several unusual properties. For one, all the different transfer RNAs of a cell have the same general shape; when purified, the heterogeneous mixture of all of a cell's transfer RNAs can form very regular crystals. The regularity of the shape of transfer RNAs makes sense. During the process of protein synthesis, two transfer RNAs attach next to each other on a ribosome, and a peptide bond forms between their amino acids. Thus, any two transfer RNAs must have the same general dimensions as well as similar structures so that they can be recognized and positioned correctly at the ribosome.

An obvious feature of the transfer RNA in figure 10.17 is that it has **unusual bases.** When this transfer RNA is originally transcribed from DNA, it is about 50% longer than the final eighty nucleotides. In fact, some transcripts contain two copies of the same transfer RNA, or sometimes several different transfer RNA genes are part of the same transcript (see fig. 10.15). The original transcription of transfer RNAs is completely regular: It does not involve unusual bases. The transcript is then processed down to the final size of a transfer RNA by various nucleases that remove trailing and leading pieces of RNA. In eukaryotes, a CCA sequence of nucleotides is added at the 3′ end by a nucleotidyl

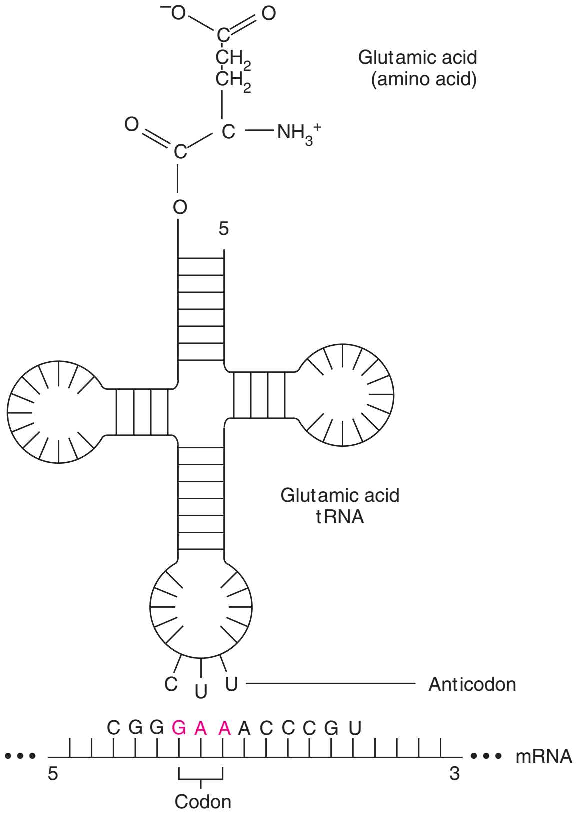

Glutamic acid
(amino acid)
$NH_3^+$
5
Glutamic acid
tRNA
C U U
Anticodon
C G G G A A A C C C G U
mRNA
5
3
Codon

**Figure 10.16** Specificity of the genetic code manifests itself in the transfer RNA, in which a particular anticodon is associated with a particular amino acid. In this case, glutamic acid is attached to its proper transfer RNA, which has the anticodon CUU.

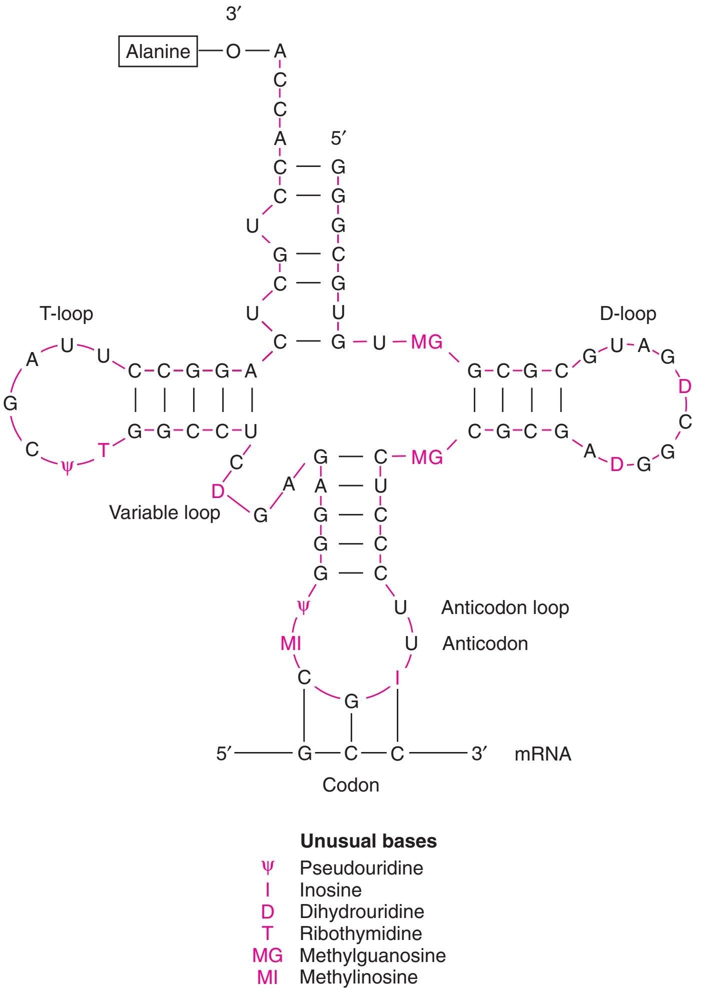

3′
Alanine
5′
T-loop
D-loop
Variable loop
Anticodon loop
Anticodon
5′
3′
mRNA
Codon
Unusual bases
ψ Pseudouridine
I Inosine
D Dihydrouridine
T Ribothymidine
MG Methylguanosine
MI Methylinosine

**Figure 10.17** Structure and sequences of alanine transfer RNA in yeast. Note the modified bases in the loops. The anticodon of the transfer RNA is shown paired with its complementary codon in the DNA. (Source: Data from R. W. Holley, et al., "Structure of a ribonucleic acid," *Science,* 147:1462–65, 1965.)

transferase enzyme. Then the transfer RNA is further modified, frequently by the addition of methyl groups to the bases already in the RNA (fig. 10.18). Presumably, these unusual bases disrupt normal base pairing and are in part responsible for the loops the unpaired bases form (see fig. 10.17).

## Transfer RNA Loops

It is believed that the first loop on the 3′ side (the T- or T-ψ-C-loop) is involved in making the transfer RNA recognizable to the ribosome. The ribosome must hold each transfer RNA in the proper orientation to check the complementarity of the anticodon of the transfer RNA and the codon of the messenger RNA. The center loop of transfer RNA is the anticodon loop. The aminoacyl-tRNA synthetases seem to recognize many points all over the transfer RNA molecule (see chapter 11).

The amino acid is attached to the ACC sequence on the 3′ end of the transfer RNA. The ribosome-binding loop on all transfer RNAs has the T-ψ-C-G sequence. The anticodon on all is bounded by uracil on the 5′ side and a purine on the 3′ side. Thus, there is a good deal of general similarity among all the transfer RNAs, consistent with the fact that they all enter protein synthesis in the same way. The actual shape of the functional transfer RNA in the cell is not an open cloverleaf, as shown in figure 10.17; rather, the whole molecule exhibits helical twisting due to pairing of complementary regions (fig. 10.19).

Earlier we considered a rough definition of a gene as a length of DNA that codes for one protein. But we have just encountered an inconsistency—genes code for both transfer RNAs and ribosomal RNAs, yet neither is eventu-

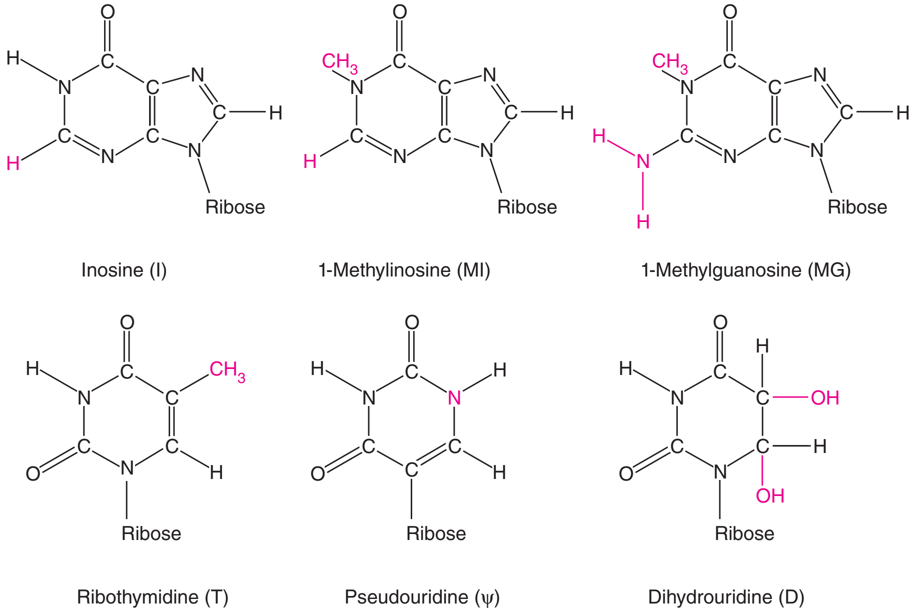

Ribose
Inosine (I)
1-Methylinosine (MI)
1-Methylguanosine (MG)
Ribothymidine (T)
Pseudouridine (ψ)
Dihydrouridine (D)

**Figure 10.18** Structures of the modified bases found in alanine transfer RNA of yeast. The various modifications of normal bases are shown in *red.*

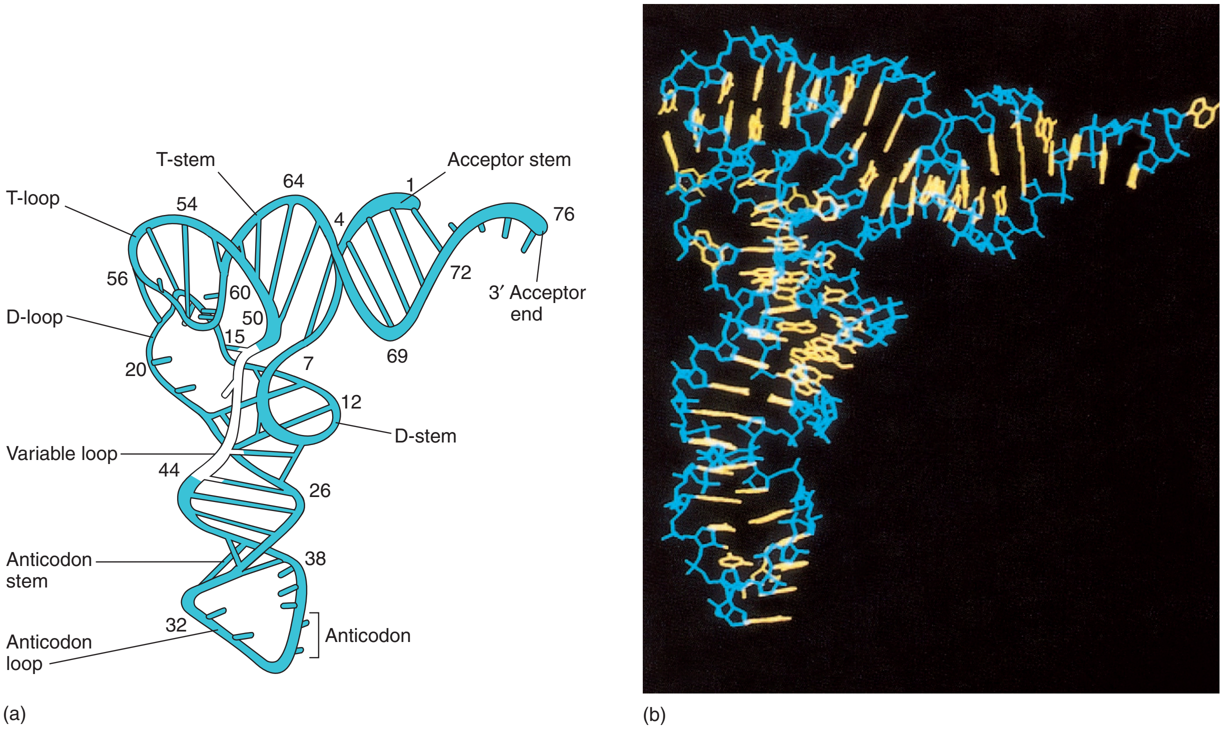

T-stem
Acceptor stem
T-loop
54
64
1
76
4
72
56
3′ Acceptor end
60
50
D-loop
15
69
7
20
12
D-stem
Variable loop
44
26
Anticodon stem
38
32
Anticodon
Anticodon loop
(a)
(b)

**Figure 10.19** Structure of yeast phenylalanine transfer RNA. (*a*) A diagram showing coiling of the sugar-phosphate backbone. (*b*) A molecular model with bases in *yellow* and backbone in *blue.* The two parts of the figure (*a* and *b*) are in the same orientation.

([*b*] Courtesy of Alexander Rich.)

ally translated into a protein. Their transcripts function as final products without ever being translated. Thus, transfer RNA and ribosomal RNA are the major exceptions to the general rule that a gene codes for a protein.

## EUKARYOTIC DNA TRANSCRIPTION

### *The Nucleolus in Eukaryotes*

Eukaryotes have four segments of ribosomal RNA in the ribosome, compared with three in prokaryotes. The smaller ribosomal subunit has an 18S piece of RNA, and the larger subunit has 5S, 5.8S, and 28S segments. All but the 5S ribosomal RNA section are transcribed as part of the same piece of RNA. However, eukaryotic cells have many copies of these ribosomal RNA genes, depending on the species. For example, the fruit fly, *Drosophila melanogaster,* has about 130 copies of the DNA region that the larger segments of ribosomal RNA are transcribed from. These regions occur in tandem on the sex (X and Y) chromosomes and are known collectively as the nucleolar organizer (see chapter 3). The smallest ribosomal RNA subunit is also produced from a duplicated gene, but at a different point in the genome. For example, in *D. melanogaster,* the 5S subunit is produced on chromosome 2.

Eukaryotes—unlike prokaryotes, which have only one RNA polymerase—have three RNA polymerases. Eukaryotic RNA polymerase I (or polymerase A) transcribes only the nucleolar organizer DNA. RNA polymerase II (or polymerase B) transcribes most genes. RNA polymerase III (or polymerase C) transcribes small genes, primarily the 5S ribosomal RNA gene and transfer RNA genes (table 10.3). In addition, mitochondria, chloroplasts, and some phages have other RNA polymerases.

**Table 10.3** **Prokaryotic and Eukaryotic RNA Polymerases**

| Enzyme | Function |
|---|---|
| *Prokaryotic* | |
| RNA polymerase | Transcribes DNA template |
| Primase | Primer synthesis during DNA replication |
| *Eukaryotic* | |
| RNA polymerase I | Transcribes nucleolar organizer |
| RNA polymerase II | Transcribes most genes |
| RNA polymerase III | Transcribes 5S rRNA and tRNA genes |
| Primase | Primer synthesis during DNA replication |

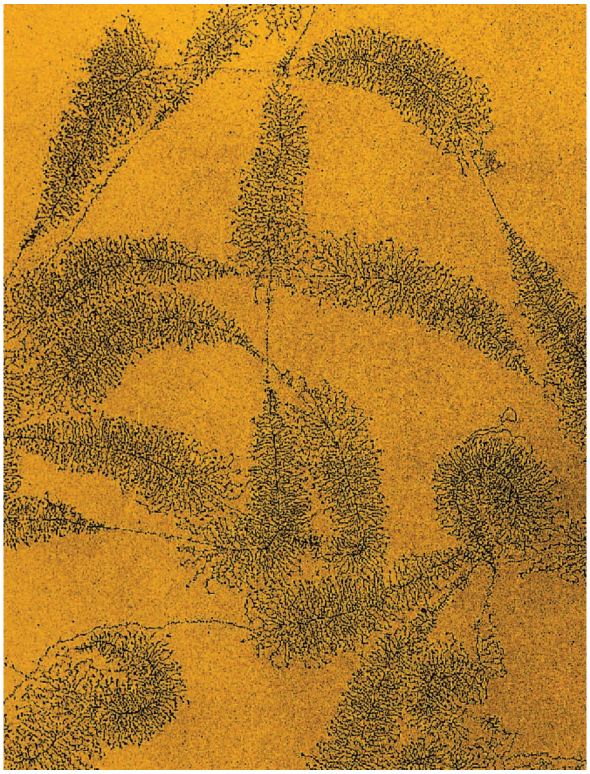


**Figure 10.20** Transcription in the nucleolus of the newt, *Triturus.* Tandem repeats of the large ribosomal RNA genes are being transcribed. The polarity of the process (progressing from small to large transcripts), as well as the spacer DNA (*thin lines* between transcribing areas), is clearly visible. Magnification 18,000×. (© O. L. Miller, B. R. Beatty, D. W. Fawcett/Visuals Unlimited.)

At the nucleolar organizer, the nucleolus forms the familiar dark blob found in eukaryotic nuclei. The nucleolus is the place where ribosomes are assembled. The various ribosomal proteins that have been manufactured in the cytoplasm migrate to the nucleus and eventually to the nucleolus, where, with the final forms of the ribosomal RNAs, they are assembled into ribosomes.

In the nucleolar organizer, an untranscribed region of spacer DNA separates each repeat of the large ribosomal RNA gene. This is shown in figure 10.20 and diagrammed in figure 10.21. In the electron micrograph in figure 10.20, the polarity of transcription is evident from the short RNA at one end of the transcribing segment and the long RNA at the other end, with a uniform gradation between. Notice that many RNA polymerases are transcribing each region at the same time. The regions between the transcribed DNA segments are the spacer DNA regions.

Like transfer RNAs, ribosomal RNAs are also modified: some uridines are converted to pseudouridines, and some ribose sugars are methylated. These conversions take place in the nucleolus, orchestrated by particles composed of small RNA segments and protein. The RNA segments are referred to as **small nucleolar RNAs (snoRNAs)** and, when combined with protein, are referred to as **small nucleolar ribonucleoprotein particles (snoRNPs).** Each different snoRNP has a snoRNA that is complementary to the regions surrounding the nucleotide to be modified. Thus, sites for modification are chosen based on complementarity to a snoRNA, which then somehow directs the modification to take place.

## *Differences Between Eukaryotic and Prokaryotic Transcription*

Although all aspects of transcription differ to some extent between prokaryotes and eukaryotes, we will look at two major differences here: the coupling of transcription and translation that is possible in prokaryotes, and the extensive **posttranscriptional modifications** that occur in eukaryotic messenger RNA. In *E. coli,* translation of the newly transcribed messenger RNA into a protein can take place before transcription is complete (fig. 10.22). The messenger RNA is synthesized in the 5′ → 3′ direction, and it is

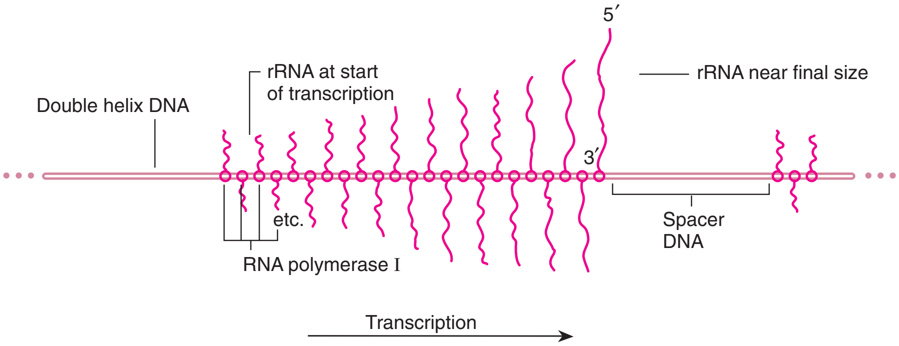

5′
rRNA at start of transcription
rRNA near final size
Double helix DNA
3′
etc.
Spacer DNA
RNA polymerase I
Transcription

**Figure 10.21** Details of the transcription of the large ribosomal RNA genes shown in figure 10.20. Note the polarity of the process and the spacer DNA, as seen in figure 10.20.

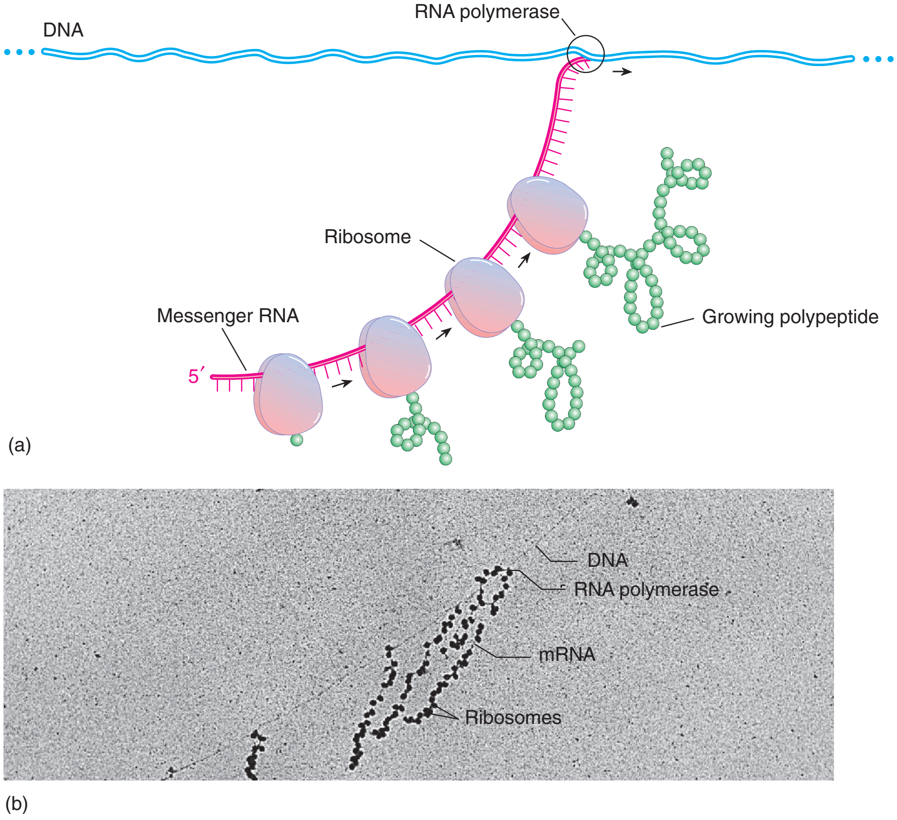

RNA polymerase
DNA
Ribosome
Messenger RNA
Growing polypeptide
5′
(a)
DNA
RNA polymerase
mRNA
Ribosomes
(b)

**Figure 10.22** (*a*) In prokaryotes, translation of messenger RNA by ribosomes begins before transcription is complete. Ribosomes attach to the growing mRNA strand when the 5′ end becomes accessible. They then move along the RNA as it elongates. When the first ribosome moves from the 5′ end, a second ribosome can attach, and so on. (*b*) Electron micrograph of events diagrammed in (*a*). The growing polypeptides cannot be seen in this preparation. Magnification 44,000×.
([*b*] Courtesy of O. L. Miller, Jr.)

near the 5′ end that translation begins. As soon as the 5′ end of the RNA is available, a ribosome can attach to the messenger RNA and move along it in the 5′ → 3′ direction, lengthening the growing polypeptide as it moves. When the first ribosome moves away from the 5′ end of the transcript, a second ribosome can attach and begin translation. These processes are repetitive, as electron micrographs (fig. 10.22*b*) clearly show. In eukaryotes, however, messenger RNA is synthesized in the nucleus, but protein synthesis takes place in the cytoplasm. (This regional division of labor is not present in *E. coli* because, among other reasons, the bacterium has no nucleus.) Before a eukaryotic messenger RNA leaves the nucleus, it is highly modified by processes that generally do not occur in prokaryotes.

## Promoters

Eukaryotic promoters are somewhat similar to prokaryotic promoters; both are regions of DNA at the beginnings of genes with signals that allow RNA polymerase to attach and begin transcription. In eukaryotes, however, more proteins are involved in promoter recognition, and more proteins are involved in the control of transcription, many recognizing signals thousands of base pairs away. We discuss these control processes in eukaryotes in chapter 16.

All three eukaryotic RNA polymerases (I, II, and III) recognize a seven-base sequence, TATAAAA, located at about −25 on the promoter DNA. It is similar to the −10 sequence in prokaryotes and is called the **TATA box** (or **Hogness box** after its discoverer, D. Hogness). Since RNA polymerase II transcribes most of the genes in eukaryotes, we turn our attention specifically to it.

Among the large number of promoters that have been sequenced, a few lack the TATA box, yet are still transcribed. Transcription initiation in these promoters appears to be controlled by a CT-rich area, called the **initiator element (Inr),** at +1 of the transcript (close to the transcription start site), coupled with a **downstream promoter element (DPE)** at about +28 to +34 of the transcript. In TATA-less promoters, a protein called TFIID requires both these elements to bind. The initiator element has a consensus sequence of TCA(G or T)T(T or C), and the downstream promoter element has the consensus sequence of (A or G)G(A or T)CGTG. We will concentrate on RNA polymerase II genes with TATA boxes.

Yeast RNA polymerase II is a protein of twelve subunits. This enzyme cannot locate promoters or attach to DNA in a stable fashion. To attach at the beginnings of genes, RNA polymerase II must interact with several proteins called **general transcription factors.** In eukaryotes, general transcription factors are named after the polymerase they work with. Thus, the transcription factor that recognizes the TATA box for polymerase II genes is called TFIID (D being the fourth letter of the alphabet for the fourth transcription factor so named). TFIID is composed of one subunit that recognizes the TATA sequence, called **TATA-binding protein (TBP),** and up to a dozen other proteins called **TBP-associated factors (TAFs),** which recognize the initiator element, when present, and aid in regulating transcription. TFIID is, in essence, similar to the sigma factors of prokaryotic RNA polymerase. One interesting aspect of the binding of TBP is that it causes a significant bending and opening of the DNA (fig. 10.23). This bending may be an important signal for other binding proteins.

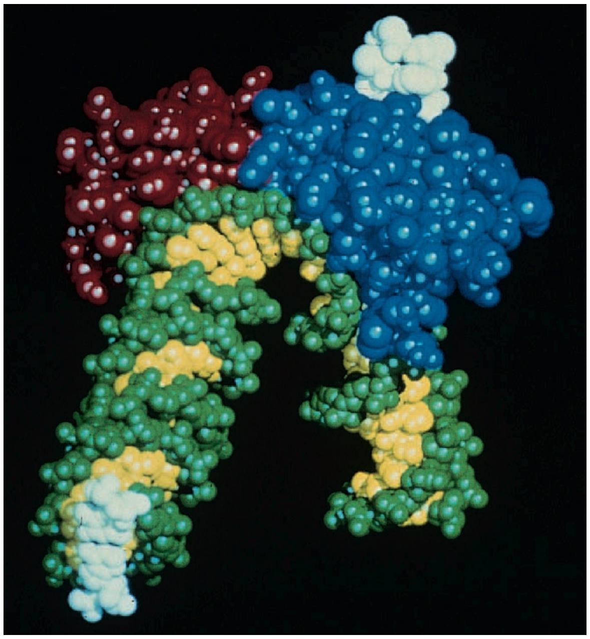


**Figure 10.23** Molecular space-filling model of a yeast TATA-binding protein attached to a TATA box on the DNA. The DNA sugar-phosphate backbone is *green* and the bases are *yellow.* The protein has twofold symmetry (*red* and *blue*). Note the bending of the DNA through 80 degrees, which also opens up the minor groove of the DNA. The *upper white* atoms are the N-terminus of the TATA-binding protein; the *lower white* atoms are the first base pair at which transcription begins. (Courtesy of J. L. Kim and S. K. Burley. From J. L. Kim, J. H. Geiger, S. Hahn, and P. B. Sigler, "Crystal Structure of a Yeast TBP TATA-box Complex." *Nature* 365 (6446): 520–27, Oct. 7, 1993. © Macmillan Magazines, Ltd. Figure adapted from the work of S. K. Burley.)

Once TFIID binds to the TATA box, a cascade of recruitment (binding) of other transcription factors takes place. Transcription factors IIA, IIB, and IIF bind, as does RNA polymerase II in an unphosphorylated state. Then transcription factors IIE and IIH bind, forming a **preinitiation complex (PIC),** equivalent to the *E. coli* holoenzyme (fig. 10.24*a*). The RNA polymerase II is then phosphorylated, presumably by TFIIH, which is a kinase; at this point, most of the transcription factors drop off, leaving the **elongation complex,** which carries out a

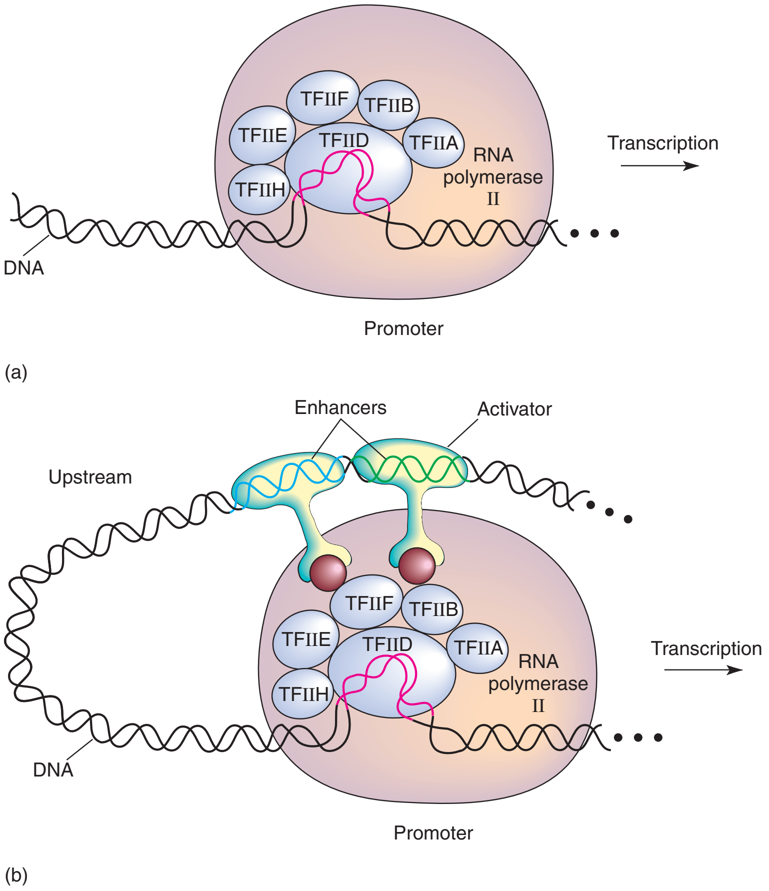


**Figure 10.24** (*a*) An RNA polymerase II preinitiation complex at a promoter. TFIID binds to the TATA box (*red*). The other transcription factors are then recruited with the polymerase. (*b*) Two activators (*yellow*) are shown bound at one end (their DNA domains) to enhancers (*blue* and *green*) upstream on the DNA. The activators are bound at their other ends (their transcriptional activation domains) to other proteins associated with the polymerase machinery. Phosphorylation of the polymerase initiates activated transcription.

**Table 10.4** **Putative Roles of the General Transcription Factors of RNA Polymerase II**

| General Transcription Factor | Function |
|---|---|
| TFIID, TBP | Recognizes TATA box |
| TFIID, TAFs | Recognizes initiator element and regulatory proteins |
| TFIIA | Stabilizes TFIID |
| TFIIB | Aids in start-site selection by RNA polymerase II |
| TFIIE | Controls TFIIH functions; enhances promoter melting |
| TFIIF | Destabilizes nonspecific interactions of RNA polymerase II and DNA |
| TFIIH | Melts promoter with helicase activity; activates RNA polymerase II with kinase activity |

*Source:* Data from R. G. Roeder, "The Role of General Initiation Factors in Transcription by RNA Polymerase II" in *Trends in Biochemical Sciences,* 21:327–35, 1996.

basal rate of transcription (fig. 10.25). TFIIH also has a role here, since it is also a helicase. Table 10.4 summarizes the postulated roles of the general transcription factors.

For activated transcription, a high level of transcription, to take place, other factors are needed that are involved in controlling which promoters are actively transcribed. These other factors are **activators** or **specific transcription factors** that bind to DNA sequences called **enhancers.** Enhancers are often hundreds or thousands of base pairs upstream from the promoter (fig. 10.24*b*).

Note that much of this information has been gathered by footprinting, mutational studies, cloning and isolating the genes and proteins involved, and then reconstituting various purified combinations in the test tube. These studies are combined with kinetic research to determine which arrangements are stable, immunological research to isolate various components with antibodies, and photocrosslinking studies to determine which moities are in contact with each other.

These specific transcriptional activators have domains (regions) that recognize their specific enhancer sequences, regions that recognize proteins associated with the polymerase (general transcription factors), and regions that allow the joint attachment of other transcription factors (fig. 10.24*b*). Similar to activators and enhancers, repressors can bind to silencer regions of DNA, often far upstream of the promoters, to repress transcription. Thus, many genes are associated with numerous and complex arrangements of transcription factors, providing elaborate control of transcription (see chapter 16).

For specific transcription factors to attach to both enhancers and the polymerase machinery, possibly thousands of base pairs apart, the DNA must bend to allow them to come into the range of the polymerase. Electron micrographs clearly show this DNA bending and looping (fig. 10.26).

Although RNA polymerases I and III seem to have termination signals similar to rho-independent promoters in prokaryotes, termination of transcription of RNA polymerase II genes is more complex, coupled with further processing of the mRNA.

Before we move on, several other points merit discussion. First, unlike prokaryotic RNA polymerases, eukaryotic RNA polymerases do proofread (showing $3' \rightarrow 5'$ exonuclease activity). Second, as we will discuss in chapter 15, eukaryotic DNA is complexed with histone proteins that can interfere with transcription. In turn, part of the RNA polymerase II complex is made up of proteins that can disrupt the histones bound to the DNA.

In addition, the RNA polymerase II complex contains proteins that act as mediators between activators and the polymerase holoenzyme. This complex coordination of the initiation of transcription in eukaryotes has been termed **combinatorial control;** the huge initiation complex may contain 85 or more different polypeptides.

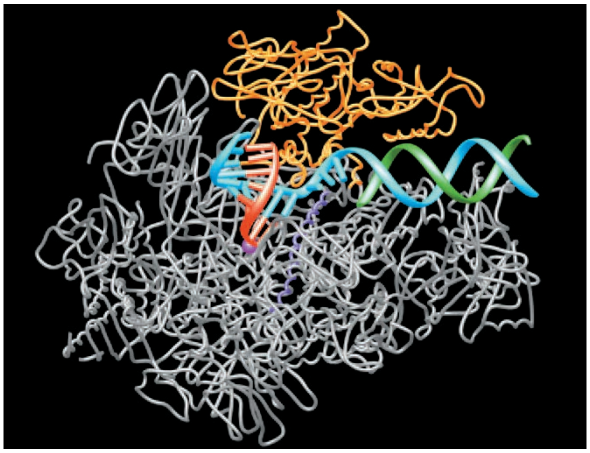
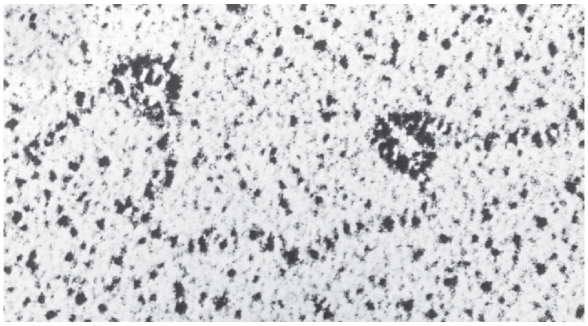
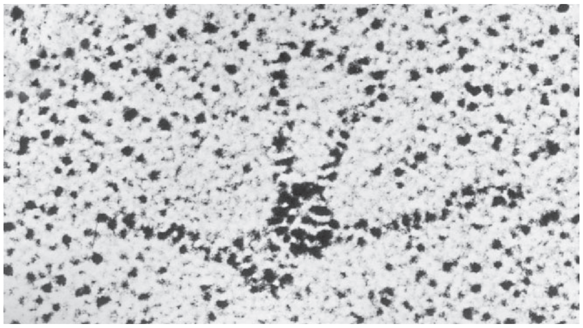


**Figure 10.25** The RNA polymerase II elongation complex with part of the protein structure removed to show the DNA and RNA within the cleft of the protein. The DNA is *blue* (template strand) and *green* (nontemplate strand) with the RNA *red*. The majority of protein is shown as *gray*; the part in *yellow* is a domain that appears to open for DNA loading and is in a closed state during elongation, thus acting as a clamp on the DNA and RNA. Closure of the clamp allows for the high stability of transcribing complexes and thus for processivity of the polymerase. The *purple* part is a helix that crosses the major cleft of the enzyme. The DNA template strand is led over this helix towards the active site. The *pink* sphere is a magnesium ion in the active site, where RNA synthesis occurs. (P. Cramer, D. A. Bushnell and R. D. Kornberg. RNA polymerase II at 2.8A resolution and A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell and R. D. Kornberg, Structure of an RNA polymerase II transcribing complex. Reprinted by permission of the authors.)



(a)



(b)

**Figure 10.26** The interaction between an activator and RNA polymerase (in this case, in prokaryotes). (*a*) In this system, the RNA polymerase of *E. coli* (the more heavily stained sphere) is controlled by an activator called NtrC (the more lightly stained sphere). The activator is bound to an enhancer, and the polymerase is bound to the promoter. (*b*) The activator has bound to the polymerase, causing a looping of the DNA. Compare with figure 10.24. (Courtesy of Sydney Kustu.)

Finally, transcription in the archaea, although under much simpler control than in the eukaryotes, resembles transcription in eukaryotes rather than prokaryotes.

The study of the details of the transcription process—its initiation, control, and termination—is one of the most active and exciting areas in modern genetics.

## Caps and Tails

Eukaryotic transcription results in a **primary transcript.** In contrast to most prokaryotic transcripts that contain information from several genes, virtually all transcripts from higher eukaryotes contain the information from just one gene. (Transcripts from several genes are found in some lower eukaryotes, such as nematode worms.) Three major changes occur in primary transcripts of RNA polymerase II before transport into the cytoplasm: modifications to the 5′ and 3′ ends and removal of intervening sequences. We refer to these changes as posttranscriptional modifications.

At the 5′ end of polymerase II transcripts, 7-methyl guanosine is added in the "wrong" direction, 5′ → 5′ (fig. 10.27). This **cap** allows the ribosome to recognize the beginning of a messenger RNA. At the other end, the 3′ end of polymerase II transcripts, a sequence of twenty to two hundred adenine-containing nucleotides, known as a **poly-A tail,** is added by the enzyme poly-A polymerase. Polyadenylation takes place after the 3′ end of the transcript is removed by a nuclease that cuts about twenty nucleotides downstream from the signal 5′-AAUAAA-3′. The tail adds stability to the molecule and aids in its transportation from the nucleus.

When messenger RNAs were first studied in eukaryotes, the messenger RNAs in the nucleus were found to be much larger than those in the cytoplasm and were called **heterogeneous nuclear mRNAs,** or **hnRNAs.** It now turns out that these were primary transcripts, RNAs that had not had any of the major posttranscriptional modifications. In essence, they were premessenger RNAs.

## Introns

Eukaryotes have segments of DNA within genes that are transcribed into RNA but never translated into protein sequences. These **intervening sequences,** or **introns,** are removed from the RNA in the nucleus before its transport into the cytoplasm (fig. 10.28). P. Sharp and his colleagues at MIT and R. Roberts, T. Broker, L. Chow, and their colleagues at the Cold Spring Harbor Laboratory first discovered introns in 1977. (Sharp and Roberts were awarded 1993 Nobel prizes for their work.) An example of a gene with introns appears in figure 10.29. The segments of the gene between introns, which are transcribed and translated—and hence exported to the cytoplasm and expressed—are termed **exons.** The results of intron removal are clear when a messenger RNA with its introns removed is hybridized with the original gene (fig. 10.30). The DNA forms double-stranded structures with the exons in RNA. The introns in DNA have nothing to pair with in the RNA, so they form single-stranded loops. Introns also occur in eukaryotic transfer RNA and ribosomal RNA genes.

For introns to be removed, the ends of the exons must be brought together and connected in a process called *splicing.* At least two types of splicing occur, although they are related: self-splicing and protein-mediated splicing.

### Self-Splicing

In 1982, Thomas Cech and his colleagues, building on the work of others, including Sidney Altman, who showed that RNA can have catalytic properties, discovered self-splicing by RNA. (Cech and Altman were awarded 1989

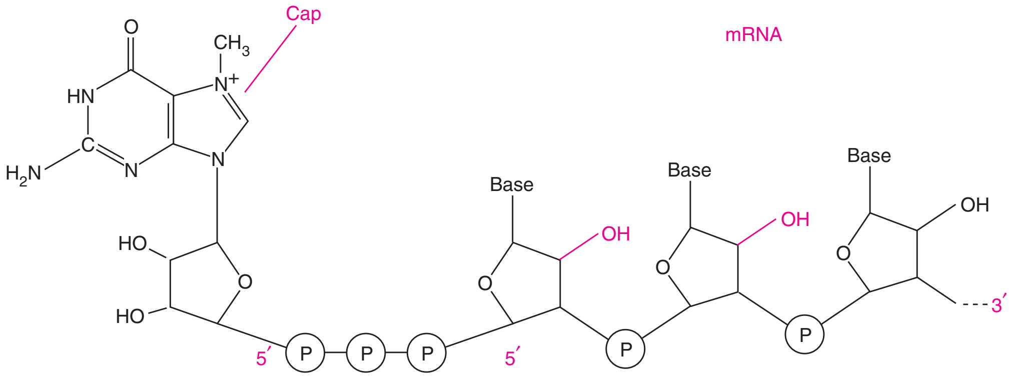
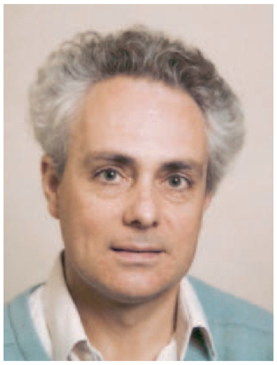


**Figure 10.27** A cap of 7-methyl guanosine is added in the "wrong" direction (5′ → 5′), to the 5′ end of eukaryotic mRNAs. In some cases, the 2′ -OH groups on the second or second and third riboses (*red*) are methylated.

Richard J. Roberts (1943–  ).
(Courtesy of Richard J. Roberts.)

Philip A. Sharp (1944–  ).
(Courtesy of Dr. Philip A. Sharp.)



Thomas Broker (1944–  ).
(Courtesy of Dr. Thomas Broker.)

Louise T. Chow (1943–  ).
(Courtesy of Dr. Louise Chow.)

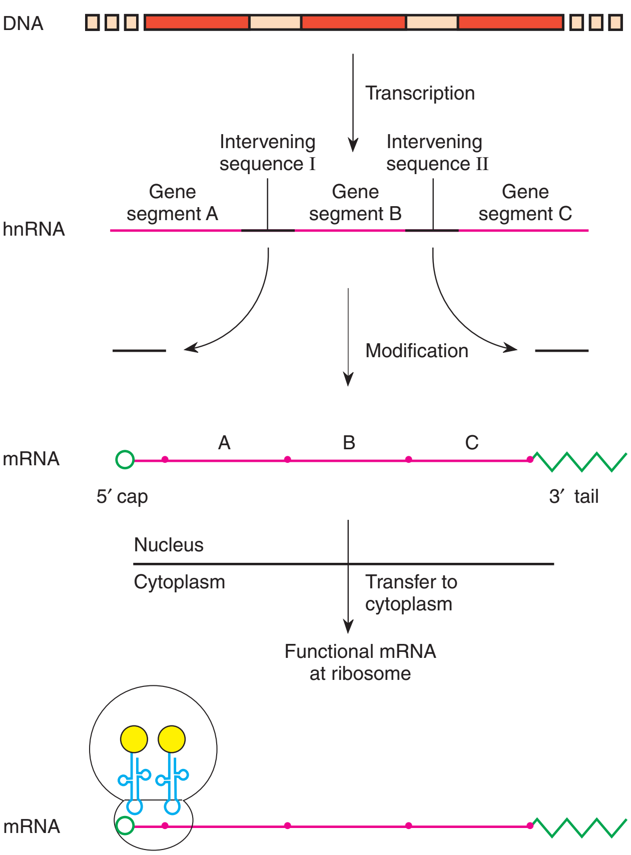


**Figure 10.28** In eukaryotic DNA, intervening sequences, or introns, are removed from the RNA in the nucleus before the mRNA is transported into the cytoplasm and translated. Other modifications consist of splicing, 5′ capping, and 3′ polyadenylation.

Nobel prizes in chemistry.) Working with an intron in the 35S ribosomal RNA precursor in the ciliated protozoan, *Tetrahymena,* Cech and his colleagues found that they could induce intron removal in vitro with no proteins present. A guanine-containing nucleotide (GMP, GDP, or GTP) had to be present. Figure 10.31 diagrams how self-splicing occurs. The intron is acting as an enzyme; we call an RNA with enzymatic properties a **ribozyme.**

During self-splicing, the U-A bond at the left (5′) side of the intron is transferred to the GTP. The U that is now unbonded displaces the G at the right (3′) side of the intron, reconnecting the RNA with a U-U connection and releasing the intron (fig. 10.31). Since all bonds are reversible transfers (transesterifications) rather than new bonds, no external energy source is required. Self-splicing introns of this type are called **group I introns.** An extensive secondary structure (RNA stem-loops) that forms is also important in intron removal (box 10.3).

Although the first enzymatic activity of the ribozyme is its own removal, its secondary structure after removal gives it the ability to further catalyze reactions (fig. 10.32). The reactions that ribozymes catalyze are transesterifications and the hydrolysis reaction of splitting an RNA molecule into two parts. Ribozymes can also perform other functions, including peptide bond formation, covered in chapter 11. Currently, at least seven different classes of ribozymes are known, based on their enzymatic properties. A ribozyme that can split other RNAs and that occurs in small plant pathogens is called a **hammerhead ribozyme** (fig. 10.33) because of its shape. Because these RNA molecules are small, they have the potential to be modified in the laboratory for specific purposes related to clinical treatment and further study of RNA processing.

Self-splicing has also been found in genes in the mitochondria of yeast. These introns are referred to as **group II introns** because they use a different mechanism of splicing that does not require an external nucleotide. Instead, the first bond is transferred within the intron to an adenosine, forming a lariat structure (fig. 10.34). In order for the lariat to form, the ribose of the adenosine must make three phosphodiester bonds (fig. 10.35).

```
              10        20        30        40        50        60        70     cAp        90
   0  GGCCAATCTGCTCACACAGGATAGAGAGGGCAGGAGCCAGGCAGAGCATATAAGGTGAGGTAGGATCAGTTGCTCCTCACATTTGCTTCTGACATAGTTG

 100  TGTTGACTCACAACCCCAGAAACAGACATCATGGTGCACCTGACTGATGCTGAGAAGGCTGCTGTCTCTTGCCTGTGGGGAAAGGTGAACTCCATGAAG
                                     MetValHisLeuThrAspAlaGluLysAlaAlaValSerCysLeuTrpGlyLysValAsnSerAspGluV
 200  TTGGTGGTGAGGCCCTGGGCAGGTTGGTATCCAGGTTACAAGGCAGCTCAAGAAGAAGTTGGGTGCTTGGAGACAGAGGTCTGCTTTCCAGCAGACAC
      alGlyGlyGluAlaLeuGlyArg  30
 300  TAACTTTCAGTGTCCCCTGTCTATGTTTCCCTTTTTAGGCTGCTGGTTGTCTACCCTTGGACCCAGCGGTACTTTGATAGCTTTGGAGACCTATCCTCTG
                                          31  LeuLeuValValTyrProTrpThrGlnArgTyrPheAspSerLeuLysGlyTh
 400  CCTCTGCTATCATGGGTAATGCCAAAGTGAAGGCCCATGGCAAGAAGGTGATAACTGCCTTTAACGATGGCCTGAATCACTTGGACAGCCTCAAGGGCAC
      laSerAlaIleMetGlyAsnAlaLysValLysAlaHisGlyLysLysValIleThrAlaPheAsnAspGlyLeuAsnHisLeuAspSerLeuLysGlyTh
 500  CTTTGCCAGCCTCAGTGAGCTCCACTGTGACAAGCTGCATGTGGATCCTGAGAACTTCAGGGTGAGTCTGATGGGCACCTCCTGGGTTTCCTTCCCCTGC
      rPheAlaSerLeuSerGluLeuHisCysAspLysLeuHisValAspProGluAsnPheArg  104
 600  TATTCTGCTCAACCTTCCTATCAGAAAAAAAGGGGAAGCGATTCTAGGGAGCAGTCTCCATGACTGTGTGTGGAGTGTTGACAAGAGTTCGGATATTTTA

 700  TTCTCTACTCAGAATTGCTGCTCCCCCTCACTCTGTTCTGTGTTGTCATTTCCTCTTTCTTTGGTAAGCTTTTTAATTTCCAGTTGCATTTTACTAAATT

 800  AATTAAGCTGGTTATTTACTTCCCATCCTGATATCAGCTTCCCCTCCTCCTTTCCTCCCAGTCCTTCTCTCTCTCCTCTCTCTTTCTCTAATCCTTTCCT

 900  TTCCCTCAGTTCATTCTCTCTTGATCTACGTTTGTTTGTCTTTTTAAATATTGCCTTGTAACTTGCTCAGAGGACAAGGAAGATATGTCCCTGTTTCTTC

1000  TCATAGCTCAAGAATAGTAGCATAATTGGCTTTTATGCAGGGTGACAGGGGAAGAATATATTTTACATATAAATTCTGTTTGACATAGGATTCTTGTGGT

1100  GGTTTGTCCAGTTTAAGGTTGCAAACAAATGTCTTTGTAAATAAGCCTGCAGGTATCTGGTATTTTTGCTCTACAGTTATGTTGATGGTTCTTCCATATT

1200  CCCACAGCTCCTGGGCAATATGATCGTGATTGTGCTGGGCCACCTTGGCAAGGATTTCACCCCCGCTGCACAGGCTGCCTTCCAGAAGGTGGTGGCT
         105  LeuLeuGlyAsnMetIleValIleValLeuGlyHisHisLeuGlyLysAspPheThrProAlaAlaPheGlnLysValValAla
1300  GGAGTGGCCACTGCCTTGGCTCACAAGTACCACTAAACCCCCTTTCCTGCTCTTGCCTGTGAACAATGGTTAATTGTTCCCAAGAGAGCATCTGTCAGTT
      GlyValAlaThrAlaLeuAlaHisLysTyrHisTer
                                            __ ___                        pA
1400  GTTGGCAAAATGATAGACATTTGAAAATCTGTCTTCTGACAAATAAAAAGCATTTATGTTCACTGCAATGATGTTTTAAATTATTTGTCTGTGTCATAGA

1500  AGGGTTTATGCTAAGTTTTCAAGATACAAAGAAGTGAGGGTTCAGGTCTCGACCTTGGGGAAATAAA
```

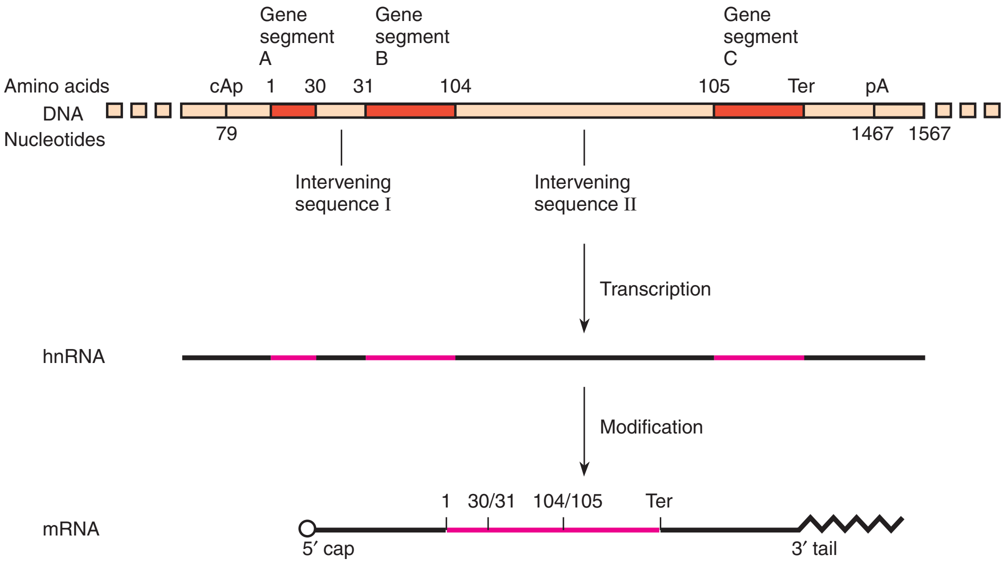


**Figure 10.29** Nucleotide sequence of the mouse β-globin major gene. The coding DNA strand is shown; cAp (position 79) indicates the start of the capped mRNA; pA indicates the start of the poly-A tail (position 1467); numbers inside the sequence are adjacent amino acid positions; Ter is the termination codon (position 1334). The three-letter abbreviations (e.g., Met, Val, His) refer to amino acids (see chapter 11). The TATA box begins at position 49. (Source: National Institutes of Health Research by David A. Konkel, et al., "The sequence of the chromosomal mouse β-globin major gene: Homologies in capping, splicing and poly (A) Sites," *Cell,* 15:1125–32, 1978.)

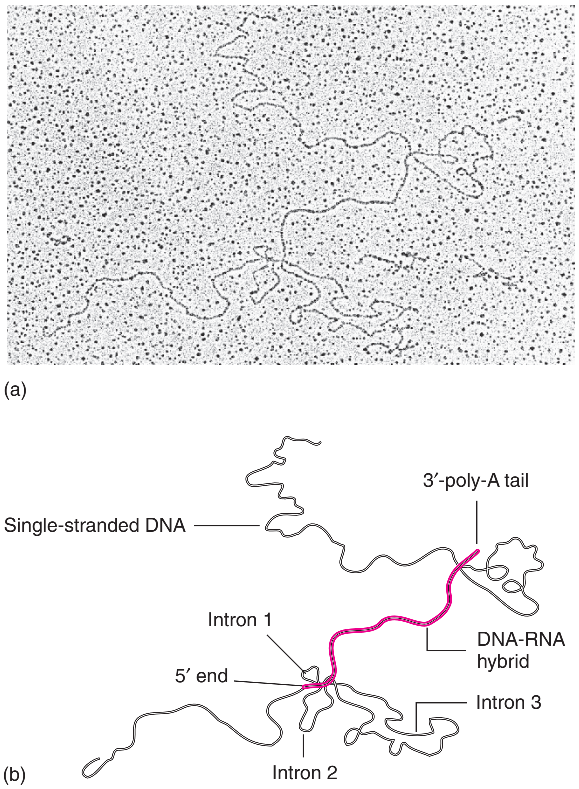
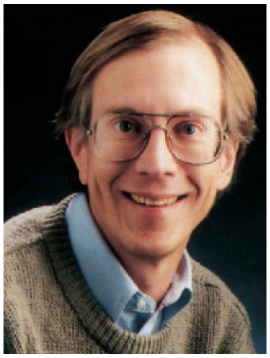


**Figure 10.30** The mRNA of adenovirus hybridized with its DNA. Three introns are visible as single-stranded DNA loops. They form single-stranded loops because they have nothing in the RNA molecule to hybridize with. Also visible is the poly-A tail of the mRNA. (*a*) Electron micrograph, (*b*) explanatory diagram. ([*a*] Courtesy of Louise T. Chow and Thomas Broker.)



Thomas Cech (1947– ). (Courtesy of Dr. Thomas Cech. Photo by Ken Abbott.)

Sidney Altman (1939– ). (Courtesy of Dr. Sidney Altman. Photo: Michael Marsland, Yale University Office of Public Affairs.)

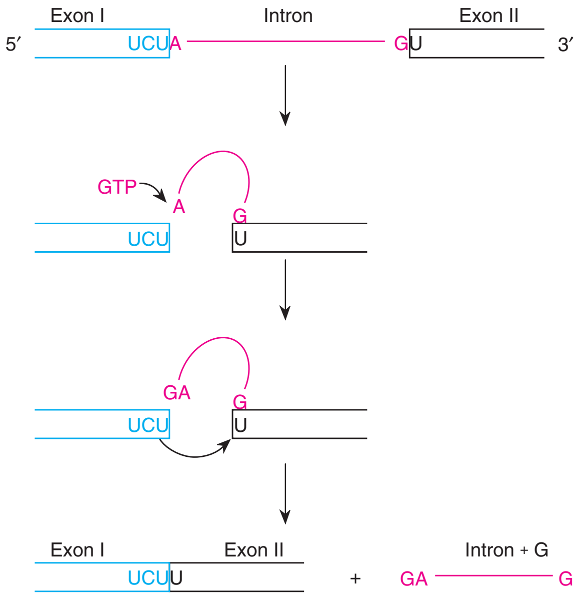


**Figure 10.31** Self-splicing of a ribosomal RNA precursor in *Tetrahymena.* An external GTP is required. Two bond transfers produce a shortened RNA and a free intron.

### *Protein-Mediated Splicing (the Spliceosome)*

Eukaryotic nuclear messenger RNAs also have their introns removed by way of a lariat structure, just as in type II introns, but with the help of RNA-protein particles. Figure 10.36 shows consensus sequences in nuclear messenger RNA for the majority of introns. At the left (5′) side of the intron, the GU sequence is invariant, as is the AG at the right (3′) side. The right-most A of the UACUAAC sequence is the branch point of the lariat and is also invariant. (In DNA nucleotides, UACUAAC is TACTAAC; therefore, that region is sometimes referred to as the **TACTAAC box.**)

Unlike the mitochondrial group II introns, however, nuclear messenger RNAs have their introns removed with the help of a protein-RNA complex called a **spliceosome,** named by J. Abelson and E. Brody. The splicing apparatus in eukaryotic messenger RNAs consists of several components called **small nuclear ribonucleoproteins** (discovered and named by J. Steitz and colleagues), abbreviated as **snRNPs** and pronounced "snurps." Five of these particles take part in splicing, each composed of one or more proteins and a small RNA molecule; they are designated U1, U2, U4, U5, and U6. The RNA molecules range in size from 100 to 215 bases. The snRNPs and their associated proteins are located in twenty to forty small regions in the nucleus called *speckles* because of their appearance in the fluorescent microscope.

The RNAs of these particles have been sequenced, and sequencing shows they have regions of complementarity to either sites in the exons, sites in the introns, or sites in the other snRNP RNAs (table 10.5). These sequences, together with the experimental techniques of photocrosslinking and the creation of selective mutations (using techniques of site-directed mutagenesis described in chapter 13) have given us insight into the splicing mechanism. Photocrosslinking tells us which components are in contact. Mutations change pairings of components and may disrupt the structure. The change can be *rescued* (the pairing restored) by making a second change in the complementary RNA. When this happens successfully, the

Joan A. Steitz (1941– ).
(Courtesy of Dr. Joan A. Steitz.)

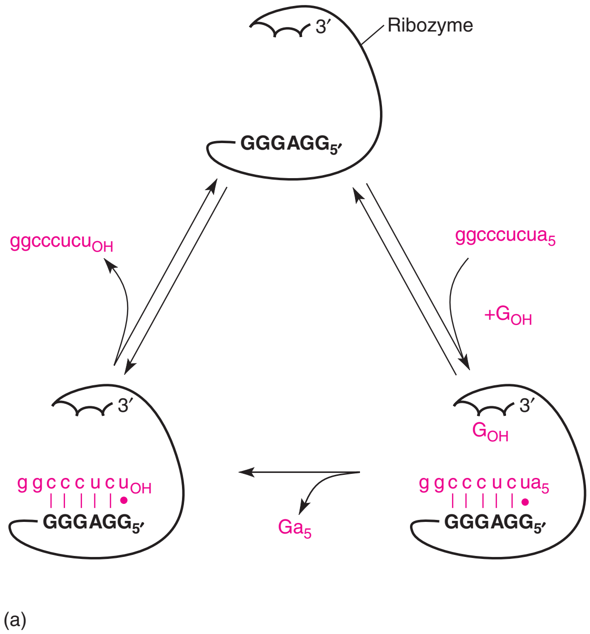

Ribozyme
3′
GGGAGG5′
ggcccucuOH
ggcccucua5
+GOH
3′
3′
GOH
g g c c c u c uOH
GGGAGG5′
g g c c c u c ua5
GGGAGG5′
Ga5
(a)

**Figure 10.32** The intron removed from the ribosomal RNA of *Tetrahymena* can catalyze the removal of the 3′ end of an RNA, diagrammed here as five AMP residues ($a_5$) from the sequence 5′-GGCCCUCUA$_5$-3′. The intron is called the *Tetrahymena* ribozyme. Any sequence can be removed from an RNA as long as there is a sequence complementary to the GGGAGG-5′ of the ribozyme to bring the RNA into position. In (*a*), the reaction needs an external guanine-containing nucleotide ($G_{OH}$); substrate nucleotides are in lowercase letters. This transesterification requires no external energy. In (*b*), the secondary structure of the ribozyme is shown. $G_{OH}$ is the site of cleavage, and the position of the G-binding site is shown. Further structure must develop to bring the G-site to the substrate. *Wavy lines* represent additional structure not shown. (Reprinted with permission from Ann Marie Pyle, et al., "RNA substrate binding site in the catalytic core of the *Tetrahymena* ribozyme," *Nature,* Volume 358, 1992. Copyright © 1992 Macmillan Magazines, Ltd.)

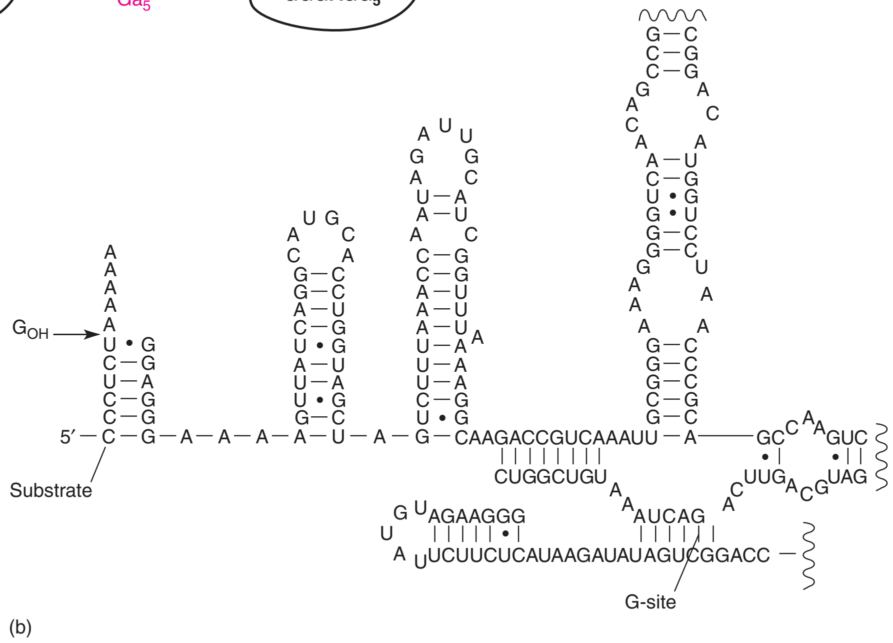

GOH
5′ – C – G – A – A – A – A – U – A – G – CAAGACCGUCAAAUU – A — GC C A A GUC
CUGGCUGU A A
UUG A C G UAG
G U AGAAGGG
U
A U UCUUCUCAUAAGAUAUAGUCGGACC
AUCAG
Substrate
G-site
(b)

**Table 10.5** The Five Small Nuclear Ribonucleoproteins (snRNPs) Involved in Nuclear Messenger RNA Intron Removal and Their RNAs

| snRNP RNA | Partial Sequence | Complementarity | Role |
|---|---|---|---|
| U1 | 3′-UCCAUUCAUA | 5′ end of intron | Recognizes and binds 5′ site of intron |
| U2 | 3′-AUGAUGU | Branch point of intron | Binds branch point of intron |
| U4 | 3′-UUGGUCGU . . . AAGGGCACGUAUUCCUU | U6 | Binds to (inactivates) U6 |
| U5 | 3′-CAUUUUCCG | Exon 1 and exon 2 | Binds to both exons |
| U6 | 3′-CGACUAGU . . . ACA | U2, 5′ site | Displaces U1 and binds 5′ site and U2 at branch point |

*Source:* With permission from the *Annual Review of Genetics*, Volume 28 © 1994 by Annual Reviews www.AnnualReviews.org.

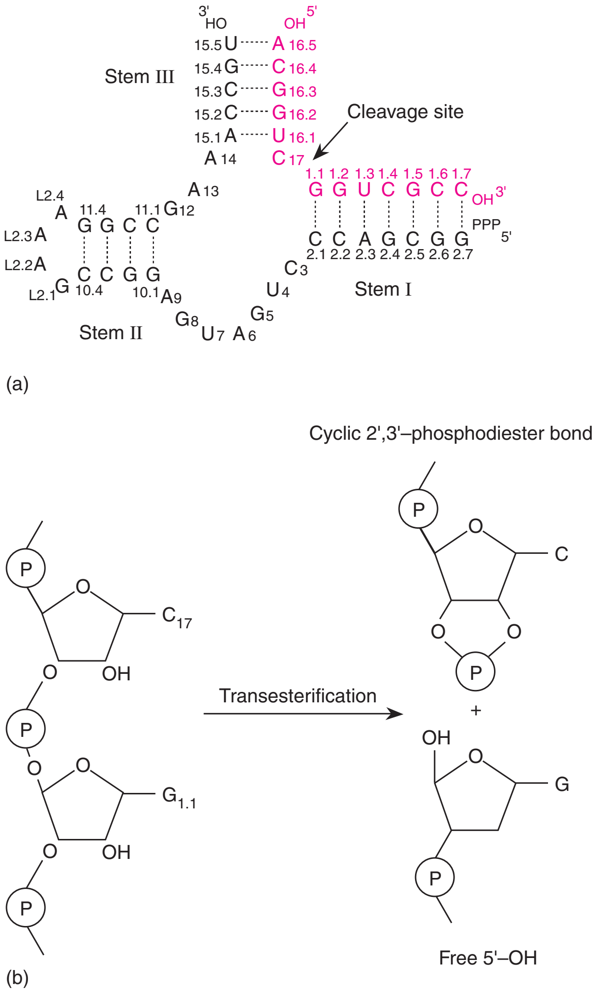

3' HO
OH 5'
15.5 U A 16.5
15.4 G C 16.4
15.3 C G 16.3
15.2 C G 16.2
15.1 A U 16.1
A 14 C 17
Stem III
Cleavage site
A 13
L2.4
A
11.4
11.1 G12
G G C C
L2.3 A
L2.2 A
C C G G
L2.1 G 10.4
10.1 A9
Stem II
G8 U7 A6
G5
U4
C3
1.1 1.2 1.3 1.4 1.5 1.6 1.7
G G U C G C C OH 3'
PPP 5'
C C A G C G G
2.1 2.2 2.3 2.4 2.5 2.6 2.7
Stem I
(a)
Cyclic 2',3'–phosphodiester bond
P
O
C
O O
P
+
OH
G
P
Free 5'–OH
Transesterification
P
O
C17
O OH
P
O
O
G1.1
O OH
P
(b)

**Figure 10.33** The hammerhead ribozyme, first seen in the RNAs of certain viruses (stems I, II, and III). (*a*) The cleavage point of the substrate (*red*) is shown using original sequence numbering, relating to the three stems of the hammerhead-shaped structure. (*b*) The cleavage, a transesterification, creates a cyclic 2′,3′-phosphodiester bond and a free 5′-OH.
(Reprinted with permission from *Nature,* Vol. 372, Heinz W. Pley et al., "Three Dimensional Structure of a Hammerhead Ribozyme." Copyright © 1994 Macmillan Magazines Limited.)

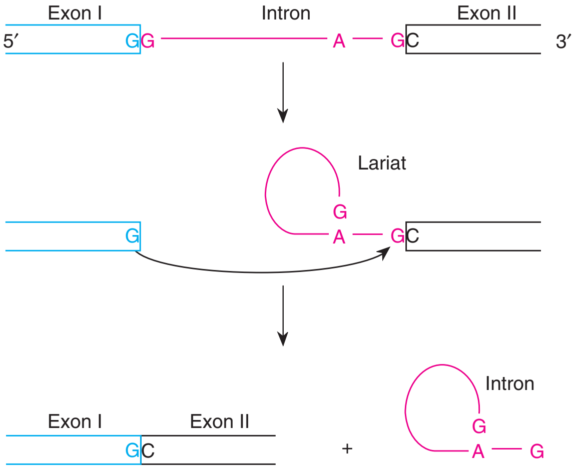


**Figure 10.34** Self-splicing of a group II intron results in a lariat configuration of the released intron. No external GTP is required since the first bond transfer takes place with an internal nucleotide, forming the loop of the lariat. A second bond transfer releases the lariat-shaped intron.

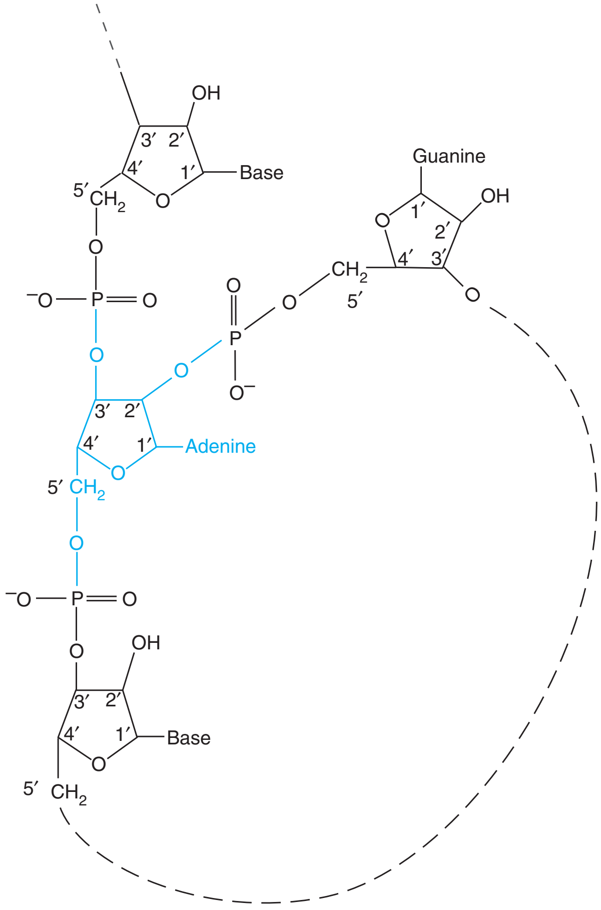


**Figure 10.35** The lariat branch point (see fig. 10.34), formed during removal of a group II intron, occurs as three phosphodiester bonds form at the same ribose sugar. (The lariat loop is formed by the 2′-phosphodiester bond.)

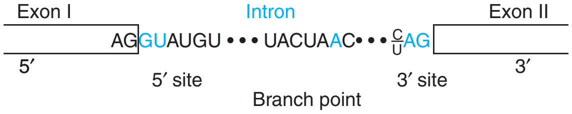


**Figure 10.36** Consensus sequences of nuclear introns, showing the 5′ and 3′ sites and the branch point. Letters in *blue* (GU, A, AG) represent invariant bases. The last A (in *blue*) of the UACUAAC sequence is the lariat branch point.

presumed pairing is then confirmed. For example, if an A-U base pair occurs between two pieces of RNA, changing the A to a C disrupts the pairing. However, if the U is converted to a G, the pairing is restored (complementary A-U bases are converted to complementary C-G bases via a noncomplementary C-U intermediate). From these techniques, we believe that the following sequence of events takes place.

First, the U1 snRNP binds at the 5′ site of the intron and the U2 snRNP binds at the branch point (fig. 10.37). The U4, U5, and U6 snRNPs form a single particle. The U4 snRNP releases, freeing the U6 snRNP to bind to the 5′ site, displacing the U1 snRNP. (The U1 snRNP, with the help of other proteins, may bind at the 5′ site simply to mark it and initiate the process.) The U6 snRNP then also binds the U2 snRNP, allowing the lariat to form in the intron. The U5 snRNP binds the two exon ends together, allowing the splice to be completed as the lariat is removed.

The splicing machinery for the majority of introns also includes numerous other polypeptides called *auxiliary* and *splicing factors;* the entire splicing process requires about 50 polypeptides. A second, less common, intron, called the U12-dependent intron, with different consensus sequences, also exists. It is removed by a similar splicing process involving different snRNPs (U11, U12) as well as many components shared with the major spliceosome.

Currently, we believe the splicing out of the intron may be autocatalyzed, just as in the type II self-splicing introns. The spliceosome may have evolved to ensure control over the process, allowing different introns to splice with differing efficiencies and allowing **alternative splicing** to take place. In many eukaryotic genes, alternative paths of splicing can take place—different splice sites may be chosen or splices may be avoided entirely. Thus, a single gene can produce several different proteins, depending on splicing choice. For example, in yeast, the gene *RPL32* codes for a ribosomal protein. When this protein accumulates in excess, it somehow causes intron removal to fail. The result is a nonfunctional messenger RNA and no further RPL32

## BOX 10.3

### Experimental Methods

### Are Viroids Escaped Introns?

**Viroids** are small (less than four hundred nucleotides), single-stranded RNA circles that act as plant pathogens. They do not have protein coats. In addition, they do not seem to code for any protein. The nature of their pathogenicity is not well understood. In 1986, Gail Dinter-Gottlieb, at the University of Colorado, pointed out numerous regions of homology between viroids and group I introns, supporting proposals by Francis Crick and Theodor Diener that viroids are escaped introns.

Many group I introns form circles after they are released. The self-splicing group I intron of *Tetrahymena thermophila* is 399 bases in circular form, whereas the potato spindle tuber viroid (PSTV) is 359 bases. These similarities of size and shape prompted the search for base homologies. In figure 1, we compare

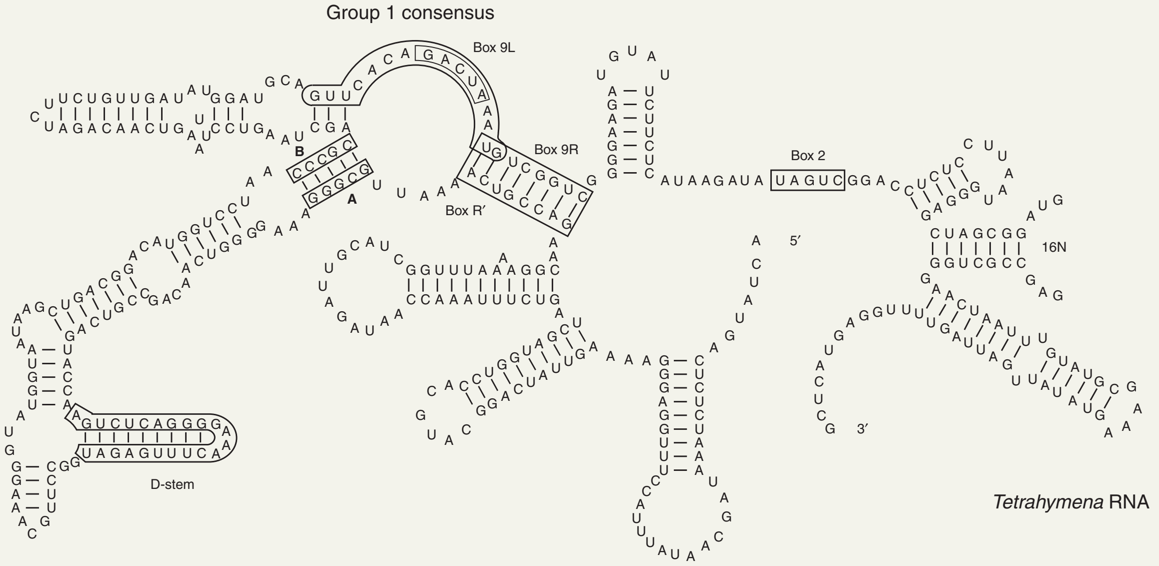


**Figure 1** Self-splicing group I *Tetrahymena* intron *(left page)* and potato spindle tuber viroid (PSTV, *right page*); 16N in the *left figure* refers to sixteen nucleotides not shown. Note the similarities around the group 1 consensus area. (From G. Dinter-Gottlieb, *Proceedings of the National Academy of Sciences,* page 6251, 1986.)

protein produced. In human beings, the gene *RBP-MS* can produce at least twelve different transcripts, depending on alternative splicing.

One other mode of protein-mediated intron removal is known. Nuclear transfer RNAs have introns that are not self-splicing but are removed by an endonuclease; the exons are subsequently joined by a ligase. Archaean bacteria seem to have this type of intron.

### Intron Function and Evolution

Since the discovery of introns, geneticists have been trying to figure out why they exist. Several views have arisen. Walter Gilbert suggested that introns separate exons (coding regions) into functional domains—that is, they separate different exons that presumably have specific tasks. In a given protein, one exon might code for a membrane-binding region, one might code for the active site of the enzyme, and one might code for ATPase activity. By recombinational mechanisms, or by excluding an exon during intron removal, **exon shuffling** would allow the rapid evolution of new proteins whose structures would be conglomerates of various functional domains. In a 1990 article in *Science,* Gilbert, with two colleagues, calculated that all proteins in eukaryotes can be accounted for by as few as one thousand to seven thousand the *Tetrahymena* intron with PSTV. Note that both have an extensive secondary structure (stem-loops) and similarities of some sequences. Most notable is the box 9L similarity. This box lies within a 16-base consensus sequence of all group I introns and has similarities to the corresponding sequence in PSTV. Note the general shape around the group I consensus region: two stems to the left and one to the right with some homologies. Note also the D-stem similarities.

These similarities strongly indicate that viroids and group I introns are related. Whether viroids are escaped introns or both evolved from a common ancestor has, as yet, not been resolved.

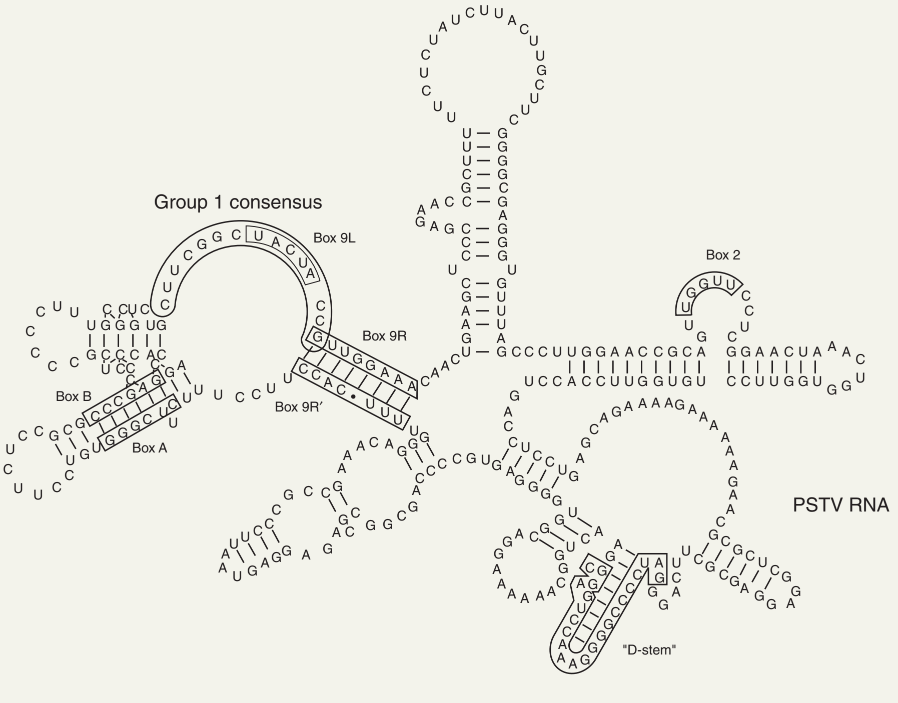


**Figure 1** (continued)

exons; all proteins may be conglomerates of this primordial number. However, this view is controversial.

J. Darnell and W. F. Doolittle have expanded Gilbert's idea of exon shuffling into the *introns-early* view. They suggest that introns arose before the first cells evolved. After eukaryotes evolved from prokaryotes, the prokaryotes lost their introns. This is supported by the evidence that, generally, prokaryotes lack introns. This view is also consistent with the opinion that the original genetic material was RNA. In this "RNA world," introns arose as part of the genetic apparatus; they were the first enzymes (ribozymes).

An alternative view is that introns arose later in evolution, after the eukaryotes split from the prokaryotes. At first, the justification for this *introns-late* view was that introns evolved late to give the organism the ability to evolve quickly to new environments by an exon-shuffling type of mechanism. However, evolutionary biologists don't accept the rationale of evolution based on future needs. An alternative explanation is that introns are actually invading "selfish DNA," DNA that can move from place to place in the genome without necessarily providing any advantage to the host organism. We call these "jumping genes" transposons and discuss them at length in chapters 14 and 16. Thus, both time frames for the development of introns—late or early—have conceptual support.

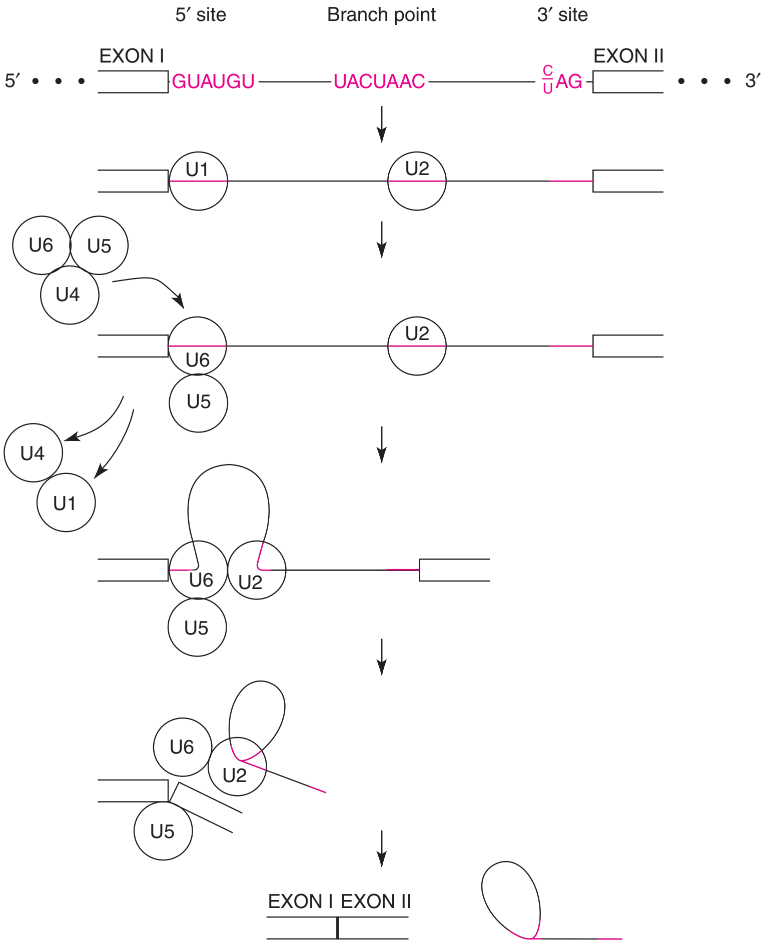


**Figure 10.37** Sequence of steps, explained in the text, in which U1, U2, U4, U5, and U6 snRNPs take part in intron removal in a nuclear RNA.

Evidence exists to support both views. Gilbert's exon-shuffling view is supported by the analysis of some genes that do indeed fit the pattern of exons coding for functional domains of a protein. (Analysis consists of DNA sequencing, RNA sequencing, and protein structural analysis.) For example, the second of three exons of the globin gene binds heme. Similarly, the human low-density lipoprotein receptor is a mosaic of exon-encoded modules shared with several other proteins. Autocatalytic properties of introns lend credence to the view that RNA was the original genetic material and that introns can move within a genome.

Additional evidence for the introns-early hypothesis includes the discovery of several introns in phage genes and introns in transfer RNA and ribosomal RNA genes in ancient bacteria (archaebacteria). Until recently, however, no introns were known in the true bacteria (eubacteria). That changed with recent work from the labs of D. Shub and J. Palmer, who independently discovered an intron in a transfer RNA gene in seven species of cyanobacteria (blue-green algae of the eubacteria). This intron was suspected to exist because it occurred in the equivalent chloroplast gene; the chloroplast evolved from an invading cyanobacterium. However, this discovery has been viewed as supporting both the introns-early and introns-late view. The introns-early supporters say this evidence confirms that introns arose before the eukaryotes-prokaryotes split. Introns-late supporters say they expect to see some introns in prokaryotes because of the mobility these bits of genetic material have.

Both the introns-early and the introns-late views may be correct. It is possible that introns arose early, were lost by the prokaryotes, which prioritized small genomes and rapid, efficient DNA replication, and later evolved to produce exon shuffling in eukaryotes.

## RNA Editing

In the last few years, several examples have arisen in which DNA sequence does not predict protein sequence. In several cases, changes in the protein occur that could have only come about by inserting or deleting nucleotides in the messenger RNA before it is translated. This insertion or deletion is almost exclusively of uridines. The process is termed **RNA editing.**

RNA editing was particularly evident in the mitochondrial proteins of a group of parasites, the trypanosomes (some of which cause African sleeping sickness); in one case, more than 50% of the nucleotides in the messenger RNA were added uridines. Uridines were also deleted from the original sequence. These parasites had another mysterious trait—the existence of minicircles and maxicircles of DNA in specialized mitochondria called *kinetoplasts.* In the average kinetoplast, there are about fifty maxicircles and about five thousand minicircles, concatenated like chain links (fig. 10.38*a*). The maxicircles contain genes for mitochondrial function (see chapter 17); as L. Simpson and his colleagues showed in 1990, both maxicircles and minicircles are templates for **guide RNA (gRNA),** RNA that guides the process of messenger editing.

The guide RNA forms a complement with the messenger RNA to be edited; however, the guide RNA has the sequence complementary to that of the *final* messenger RNA, the one with bases added. Since the bases have not yet been added, a bulge occurs in the guide RNA where the complement to be added is (fig. 10.38*b*). The messenger RNA is then cleaved opposite the bulge by an editing endonuclease. A uridylate (U) is brought into the messenger RNA as a complement to the adenine (A) with the enzyme terminal-U-transferase. An RNA ligase then closes the nick in the messenger RNA, which now has a uridylate added.

An exciting outcome of this research, aside from learning about a novel mechanism of messenger RNA processing, is the possibility of clinical rewards. Anytime there is a specialized pathway in a parasite not found in its host, it is possible to use that pathway to attack the parasite. Thus, this research might lead to new ways of combating these trypanosome parasites.

RNA editing also occurs in other species and by different mechanisms. For example, in the apolipoprotein-B (*apoB*) gene in mammals, one gene produces two forms of the protein. In one case, nucleotide 6666, a cytosine, is modified by deamination to a uracil in the messenger RNA, resulting in the termination of translation and a protein about half the normal size. RNA editing also occurs in plant mitochondria and chloroplasts in which the usual change is also a cytosine to a uracil. RNA editing is thus routinely seen in specific examples of posttranscriptional RNA modification in both animals and plants.

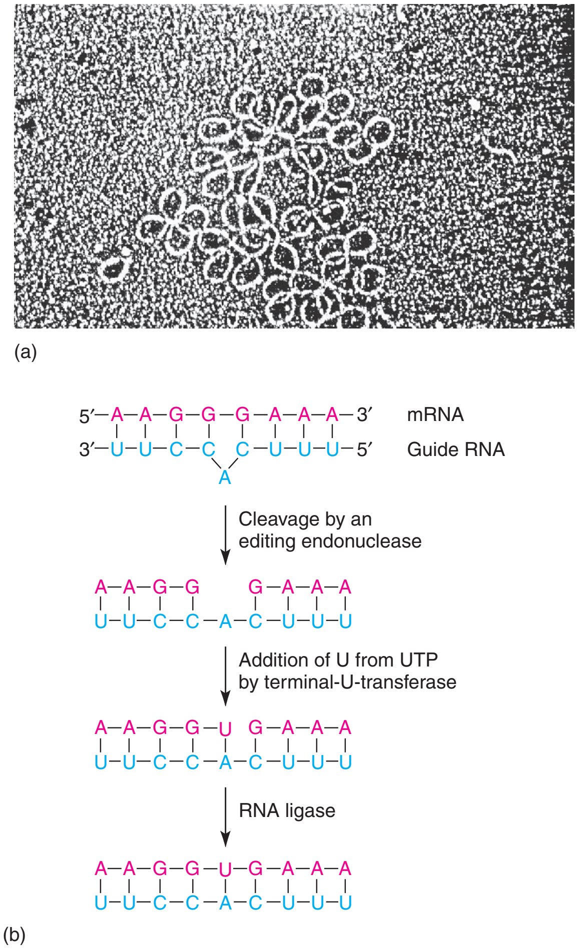


**Figure 10.38** RNA editing. (*a*) Eight hundred seventy base pair minicircles of DNA from *Leishmania tarentolae.* (*b*) Mechanism by which a guide RNA is involved in the editing of a messenger RNA. After the cycle shown, a uridine-containing nucleotide has been added to the messenger RNA. The guide RNA has the sequence complementary to the messenger RNA with the base already added. ([*a*] Courtesy of Larry Simpson.)

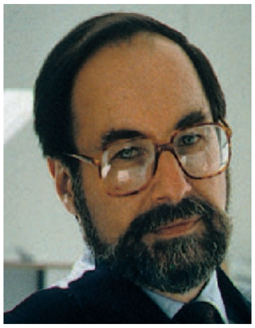
## UPDATED INFORMATION ABOUT THE FLOW OF GENETIC INFORMATION

The original description of the central dogma included three information transfers that were presumed to occur even though they had not been observed (see fig. 10.1). Since then, researchers have documented these three transfers: reverse transcription, RNA self-replication, and the direct involvement of DNA in translation (fig. 10.39).

## Reverse Transcription

First, the return arrow from RNA to DNA in figure 10.39 indicates that RNA can be a template for DNA synthesis. All RNA tumor viruses, such as Rous sarcoma virus, as well as the AIDS virus, can make an RNA-dependent DNA polymerase (often referred to as **reverse transcriptase**) that synthesizes a DNA strand complementary to the viral RNA. (H. Temin and D. Baltimore received Nobel prizes for their discovery of this polymerase enzyme.) This enzyme is involved in a tumor virus's infection of a normal cell and the transformation of that cell into a cancerous cell. When the viral RNA enters a cell, it brings reverse transcriptase with it. The enzyme synthesizes a DNA-RNA double helix, which then is enzymatically converted into a DNA-DNA double helix that can integrate into the host chromosome. After integration, the DNA is transcribed into copies of the viral RNA, which are both translated and packaged into new viral particles that are released from the cell to repeat the infection process. (We cover this material in more detail in chapters 13 and 16.)

Howard Temin (1934–1994). (Courtesy of Dr. Howard Temin. UW photo media.)



David Baltimore (1938– ). (Courtesy of Kucerea and Company/Laxenburger Strasse 58.)

## RNA Self-Replication

The second modification to the original central dogma is the verification that RNA can act as a template for its own replication. This process has been observed in a small class of phages. These **RNA phages,** such as R17, f2, MS2, and Qβ, are the simplest phages known. MS2 contains about thirty-five hundred nucleotides and codes for only three proteins: a coat protein, an attachment protein (responsible for attachment to and subsequent penetration of the host), and a subunit of the enzyme **RNA replicase.** The RNA replicase subunit combines with three of the cell's proteins to form RNA replicase, allowing the single-stranded RNA of the phage to replicate itself.

Since the new protein needed to construct the RNA replicase enzyme must be synthesized before the phage can replicate its own RNA, the phage RNA must first act as a messenger when it infects the cell. Thus, protein synthesis is taking place without a preceding transcription process. The viral genetic material, its RNA, is first used as a messenger in the process of translation and then used as a template for RNA replication.

## DNA Involvement in Translation

In the mid-1960s, B. J. McCarthy and J. J. Holland showed that under certain experimental conditions, denatured (single-stranded) DNA could bind to ribosomes and be translated into proteins. The experimental conditions usually involved the addition of antibiotics that interacted with the DNA or the ribosome. Direct translation of DNA is not known to occur naturally.

Even in our updated central dogma (fig. 10.39), no arrows originate at the protein. In other words, protein cannot self-replicate, nor can it use amino acid sequence information to reconstruct RNA or DNA. Crick has called these arrows "forbidden transfers." We know of no cellular machinery that can produce these forbidden processes. In the next chapter, we continue this discussion of protein synthesis by describing the process of translation, in which the information in messenger RNA is used to form the sequences of amino acids in proteins.

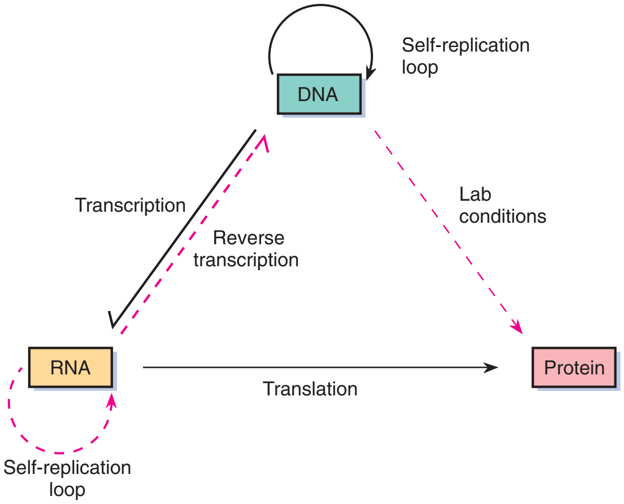


**Figure 10.39** An updated version of Crick's central dogma, showing all known paths of genetic information transfer. Paths confirmed since Crick proposed the original central dogma appear as dashed *red* lines (reverse transcription, RNA self-replication, and direct DNA translation). Direct DNA translation is known only under laboratory conditions: the process apparently does not occur naturally. There is no known information flow beginning with protein.

## SUMMARY

The central dogma is a description of how genetic information is transferred among DNA, RNA, and protein. In chapter 9, we described the DNA self-replication loop. In this chapter, we described the transcriptional process, in which DNA acts as a template for the production of RNA.

**STUDY OBJECTIVE 1:** To examine the types of RNA and their roles in gene expression 245–246, 256–260

Messenger RNA (mRNA) is a complementary copy of the DNA of a gene that carries the information of the gene to the ribosomes, where protein synthesis actually takes place. Transfer RNAs (tRNAs) transport the amino acid building blocks of proteins to the ribosome. Complementarity between the messenger RNA codon and the transfer RNA anticodon establishes the amino acid sequence in the synthesized protein ultimately specified by the gene. Ribosomal RNA (rRNA) is also involved in this process of gene-directed protein synthesis.

**STUDY OBJECTIVE 2:** To look at the process of transcription, including start and stop signals, in both prokaryotes and eukaryotes 246–256

Intracellular RNA is single-stranded, although extensive intramolecular stem-loop structures may form. At any one gene, RNA is transcribed from only one strand of the DNA double helix. The transcribing enzyme is RNA polymerase. In *E. coli,* the core enzyme, when associated with a sigma factor, becomes the holoenzyme that recognizes the transcription start signals in the promoter. Several consensus sequences define a promoter. In prokaryotes, termination of transcription requires a sequence on the DNA, called the terminator, that causes a stem-loop structure to form in the RNA. Sometimes the rho protein is required for termination (in rho-dependent, as compared with rho-independent, termination). In eukaryotes, there are three RNA polymerases. Eukaryotic genes have promoters with sequences analogous to those in prokaryotic promoters as well as enhancers that work at a distance.

The ribosome is made of two subunits, each with protein and RNA components. Transfer RNAs are charged with their particular amino acids by enzymes called aminoacyl-tRNA synthetases. Each transfer RNA has about eighty nucleotides, including several unusual bases. All transfer RNAs have similar structures and dimensions. Transfer RNAs and ribosomal RNAs are modified from their primary transcripts.

**STUDY OBJECTIVE 3:** To investigate posttranscriptional changes in eukaryotic messenger RNAs, including an analysis of intron removal 260–276

Prokaryotic messenger RNAs are transcribed with a leader before, and a trailer after, the translatable part of the gene. In prokaryotes, translation begins before transcription is completed. In eukaryotes, these processes are completely uncoupled—transcription is nuclear and translation is cytoplasmic. Eukaryotic messenger RNA is modified after transcription: a cap and tail are added, and intervening sequences (introns) are removed, before transport into the cytoplasm. Introns can be removed by self-splicing or with the aid of the spliceosome, composed of small nuclear ribonucleoproteins (snRNPs). It is not known whether introns arose early or late in evolution or what their functions are. In some organisms, such as trypanosomes, RNAs can be edited further by the addition or deletion of nucleotides under the direction of guide RNA.

The study of several RNA viruses has shown that RNA can act as a template to replicate itself and to synthesize DNA; under laboratory conditions, DNA can be translated directly into protein. These discoveries add new directions of information transfer to the central dogma.

## SOLVED PROBLEMS

**PROBLEM 1:** What would be the sequence of segments on a prokaryotic messenger RNA with more than one gene present?

***Answer:*** The transcript would have unmodified 5′ (leader) and 3′ (trailer) ends. Reading the sequence of nucleotides on the RNA, you would come across an initiation codon (AUG) and then, after perhaps nine hundred more nucleotides, a termination codon (UAA, UAG, or UGA). The nine hundred nucleotides would be those translated into the protein. Then there would be a spacer region of nucleotides, followed by another initiation codon, intervening nucleotides that are translated into amino acids, and a termination codon. This sequence of initiation codon, codons to be translated, a termination codon, and spacer RNA would be repeated for as many genes as are present in the messenger RNA.

**PROBLEM 2:** Can one nucleotide be a conserved sequence?

***Answer:*** Conserved sequences are invariant sequences of DNA or RNA recognizable to either a protein or a complementary sequence of DNA or RNA. However, in group II introns, an adenine is needed near the 3′ end of the intron for lariat formation. Thus, this single nucleotide, given its relative position in the intron and possible surrounding bases, is a conserved sequence of one.

**PROBLEM 3:** Why might *E. coli* not have a nucleolus?

***Answer:*** The nucleolus is the site of ribosomal construction in eukaryotes. It is centered at the nucleolus organizer, the tandemly repeated gene coding for the three larger pieces of ribosomal RNA. In *E. coli,* there are only five to ten copies of the ribosomal RNA gene, whereas there is usually an order of magnitude or more copies in eukaryotes. Thus, the simplest reason that a nucleolus is not visible in *E. coli* is because there are too few copies of the gene around which a nucleolus forms.

**PROBLEM 4:** If this sequence of bases represents the start of a gene on double-stranded DNA, what is the sequence of the transcribed RNA, what is its polarity, and what is the polarity of the DNA?

G C T A C G G A T T G C T G
C G A T G C C T A A C G A C

***Answer:*** Begin by writing the complementary strand to each DNA strand: C G A U G C C U A A C G A C for the top, and G C U A C G G A U U G C U G for the bottom. Now look for the start codon, AUG. It is present only in the RNA made from the top strand, so the top strand must have been transcribed. The polarity of the start codon is 5′-A U G-3′. Since transcription occurs 5′ → 3′, and since nucleic acids are antiparallel, the left end of the top DNA strand is the 3′ end.

## EXERCISES AND PROBLEMS*

### TYPES OF RNA

1. Diagram the relationships of the three types of RNA at a ribosome. Which relationships make use of complementarity?

### PROKARYOTIC DNA TRANSCRIPTION

2. How could DNA-DNA or DNA-RNA hybridization be used as a tool to construct a phylogenetic (evolutionary) tree of organisms?
3. Assume that prokaryotic RNA polymerase does not proofread. Do you expect high or low levels of error in transcription as compared with DNA replication? Why is it more important for DNA polymerase than RNA polymerase to proofread?
4. What are the transcription start and stop signals in eukaryotes and prokaryotes? How are they recognized? Can a transcriptional unit include more than one translational unit (gene)? (*See also* EUKARYOTIC DNA TRANSCRIPTION)
5. What is a consensus sequence? a conserved sequence?
6. What would the effect be on transcription if a prokaryotic cell had no sigma factors? no rho protein?
7. Draw a double helical section of prokaryotic DNA containing transcription start and stop information. Give the base sequence of the messenger RNA transcript.
8. In what ways does the transcriptional process differ in eukaryotes and prokaryotes? (*See also* EUKARYOTIC DNA TRANSCRIPTION)
9. What is a stem-loop structure? an inverted repeat? a tandem repeat? Draw a section of a DNA double helix with an inverted repeat of seven base pairs.
10. What is the function of each of the following sequences: TATAAT, TTGACA, TATA, TACTAAC? What is a Pribnow box? a Hogness box? (*See also* EUKARYOTIC DNA TRANSCRIPTION)
11. What is footprinting? How did it help define promoter sequences?
12. What are the differences between rho-dependent and rho-independent termination of transcription?
13. What are the differences between a $\sigma^{70}$ and a $\sigma^{32}$?
14. Draw a typical mature messenger RNA molecule of a prokaryote and a eukaryote. Label all regions. (*See also* EUKARYOTIC DNA TRANSCRIPTION)
15. Determine the sequence of both strands of the DNA from which this RNA was transcribed. Indicate the 5′ and 3′ ends of the DNA and, with an arrow, which strand was transcribed.

5′-C C A U C A U G A C A G A C C C U U G C U A A C G C-3′

16. The following DNA fragment was isolated from the beginning of a gene. Determine which strand is transcribed, indicate the polarity of the two DNA strands, and then give the sequence of bases in the resultant messenger RNA and its polarity.

C C C T A C G C C T T T C A G G T T
G G G A T G C G G A A A G T C C A A

---

*Answers to selected exercises and problems are on page A-11.

**17.** The following DNA fragment represents the beginning of a gene. Determine which strand is transcribed, and indicate polarity of both strands in the DNA.

ATGATTTACATCTACATTTACATT
TACTAAATGTAGATGTAAATGTAA

**18.** The following sequence of bases in a DNA molecule is transcribed into RNA:

CCAGGTATAATGCTCCAGTATGGCATGGTACTTCCGG
↑

If the T (*arrow*) is the first base transcribed, determine the sequence and polarity of bases in the RNA, and identify the Pribnow box and the initiator codon.

**19.** You have isolated a mutant that makes a temperature-sensitive rho molecule; rho functions normally at 30° C, but not at 40° C. If you grow this strain at both temperatures for a short period of time and isolate the newly synthesized RNA, what relative size RNA do you expect to find in each case?

**20.** Suppose you repeat the experiment in problem 19 and find the same size RNA made at both temperatures. Provide two possible explanations for this unexpected finding.

**21.** Why do you think most promoter regions are A-T rich? (*See also* EUKARYOTIC DNA TRANSCRIPTION)

### EUKARYOTIC DNA TRANSCRIPTION

**22.** Would introns be more or less likely than exons to accumulate mutations through evolutionary time?

**23.** What would be the effect on the final protein product if an intervening sequence were removed with an extra base? one base too few?

**24.** What is heterogeneous nuclear messenger RNA? What are small nuclear ribonucleoproteins?

**25.** What product would DNA-RNA hybridization produce in a gene with five introns? no introns? Draw these hybrid molecules.

**26.** What are the recognition signals within the majority of introns?

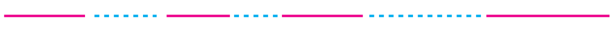
**27.** What are the differences between group I and group II introns?

**28.** Diagram ribozyme functioning in a group I intron.

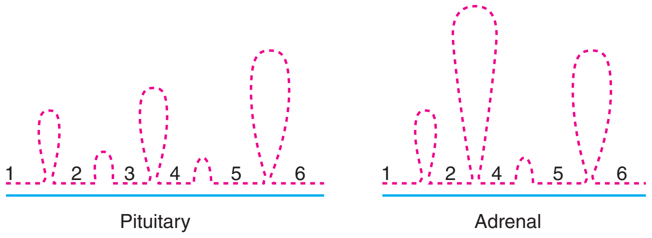
**29.** How does a spliceosome work? What are its component parts?

**30.** What is a transcriptional factor? an enhancer?

**31.** In the following drawing of a eukaryotic gene, solid red lines represent coding regions, and dashed blue lines represent introns. Draw the RNA-DNA hybrid that would result if cytoplasmic messenger RNA is hybridized to nuclear DNA.



**32.** RNA-DNA hybrids are formed by using messenger RNA for a given gene that is expressed in the pituitary and the adrenal glands. The DNA used in each case is the full-length gene. Based on the figure, provide an explanation for the different hybrid molecules. DNA is a dashed red line; RNA is a solid blue line.



**33.** Enhancers can often exert their effect from a distance; some enhancers are located thousands of bases upstream from the promoter. Propose an explanation to account for this.

### UPDATED INFORMATION ABOUT THE FLOW OF GENETIC INFORMATION

**34.** How do prions relate to the central dogma of figure 10.39?

## CRITICAL THINKING QUESTIONS

**1.** What are the upper limits to the size of a gene in eukaryotes?

**2.** Present a scenario of an activator controlling transcription in eukaryotes.

*Suggested Readings for chapter 10 are on page B-7.*

# 11

# GENE EXPRESSION

## *Translation*

## STUDY OBJECTIVES

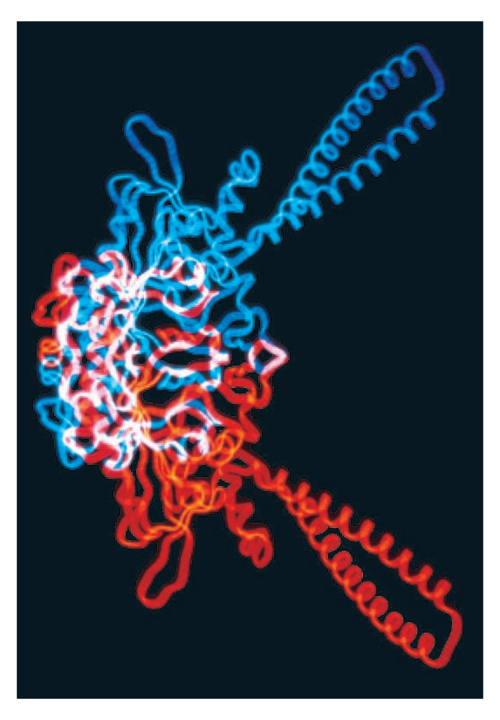
1. To study the mechanism of protein biosynthesis, in which organisms, using the information in DNA, string together amino acids to form proteins 281
2. To examine the genetic code 304

## STUDY OUTLINE

**Information Transfer** 281
- Transfer RNA 281
- Initiation Complex 288
- Elongation 292
- Termination 296
- More on the Ribosome 299
- The Signal Hypothesis 301
- The Protein-Folding Problem 303

**The Genetic Code** 304
- Triplet Nature of the Code 304
- Breaking the Code 305
- Wobble Hypothesis 307
- Universality of the Genetic Code 308
- Evolution of the Genetic Code 311

*Summary* 312
*Solved Problems* 313
*Exercises and Problems* 313
*Critical Thinking Questions* 314
**Box 11.1 Amino Acid Sequencing** 284
**Box 11.2 Antibiotics** 294



Computer-generated model of the enzyme seryl tRNA synthetase, the enzyme that charges tRNA's with the amino acid serine. (© Dr. Stephen Cusack/EMBL/SPL/Photo Researchers, Inc.)

In this chapter, we continue our discussion of gene expression, concentrating on protein biosynthesis. This process, which translates the nucleotide information in messenger RNA into amino acid sequences in proteins, is the final step of the central dogma. Nucleotide sequences in DNA are transcribed into nucleotide sequences in RNA, which are then translated into amino acid sequences in proteins.

All proteins are synthesized from only twenty naturally occurring amino acids (fig. 11.1). (There is one exception, selenocysteine, which we discuss at the end of the chapter.) These are called α-amino acids because one carbon, the α carbon, has four specific groups attached to it: an amino group, a carboxyl (acidic) group, a hydrogen, and one of the twenty different R groups (side chains), imparting the specific properties of that amino acid. (Technically, proline is termed an *imino* acid because of its structure.) Having these four groups attached imparts a property known as *chirality* on the amino acid: like left- and right-handed gloves, the mirror images cannot be superimposed. Because of optical properties, the two forms of each amino acid are referred to as D and L, in which D comes from dextrorotatory (right turning) and L comes from levorotatory (left turning). All biologically active amino acids are of the L form, and hence we need not refer to this designation. Proteins (polypeptides) are synthesized when peptide bonds form between any two amino acids (fig. 11.2). In this manner, long chains of amino acids—called *residues* when incorporated into a protein—can join, and all chains will have an amino (N-terminal) end and a carboxyl (C-terminal) end.

The sequence of polymerized amino acids determines the **primary structure** of a protein. Included in the primary structure is the formation of disulfide bridges between cysteine residues (fig. 11.3). Polypeptides can fold into several structures, the most common of which are α helices and β sheets. These folding configurations constitute the **secondary structure** of the protein. In some proteins, the folding is spontaneous; in some, it is guided by other proteins. Further folding, bringing α helices and β sheets into three-dimensional configurations, creates the **tertiary structure** of the protein (fig. 11.4). Many proteins in the active state are composed of several subunits that together make up the **quaternary structure** of the protein. Translation is the process in which the primary structure of a protein is determined from the nucleotide sequence in a messenger RNA (box 11.1).

## INFORMATION TRANSFER

Before proceeding to the details of translation, a sketch of the beginning of the process may be helpful (fig. 11.5). The ribosome with its ribosomal RNA and proteins is the site of protein synthesis. The information from the gene is in the form of messenger RNA, in which each group of three nucleotides—a codon—specifies an amino acid. The amino acids are carried to the ribosome attached to transfer RNAs, and these transfer RNAs have anticodons, three nucleotides complementary to a codon, located at the end opposite the amino acid attachment site. A peptide bond will form between the two amino acids present at the ribosome, freeing one transfer RNA (at codon 1 in fig. 11.5) and lengthening the amino acid chain attached to the second transfer RNA (at codon 2 in fig. 11.5). The messenger RNA will then move one codon with respect to the ribosome, and a new transfer RNA will attach at codon 3. This cycle is then repeated, with the polypeptide lengthening by one amino acid each time. We can begin looking at the details of translation by looking at the transfer RNAs. As before, we concentrate on the prokaryotic system, noting details about eukaryotes as appropriate.

## Transfer RNA

### Attachment of Amino Acid to Transfer RNA

The function of transfer RNA is to ensure that each amino acid incorporated into a protein corresponds to a particular codon (a group of three consecutive nucleotides) in the messenger RNA. The transfer RNA serves this function through its structure: It has an anticodon at one end and an amino acid attachment site at the other end. The "correct" amino acid, the amino acid corresponding to the anticodon, is attached to the transfer RNA by enzymes known as aminoacyl-tRNA synthetases (e.g., arginyl-tRNA synthetase, leucyl-tRNA synthetase). A transfer RNA with an amino acid attached is said to be "charged."

An aminoacyl-tRNA synthetase joins a specific amino acid to its transfer RNA in a two-stage reaction that takes place on the surface of the enzyme. In the first stage, the amino acid is activated with ATP. In the second stage of the reaction, the amino acid is attached with a high-energy bond to the 2′ or 3′ carbon of the ribose sugar at the 3′ end of the transfer RNA (fig. 11.6). In the figure, we denote high-energy bonds, bonds that liberate a lot of free energy when hydrolyzed, as "~." Thus, during the process of protein synthesis, the energy for the formation of the peptide bond will be present where it is needed, at the point of peptide bond formation.

### Component Numbers

In bacteria, there are twenty aminoacyl-tRNA synthetases, one for each amino acid. A particular enzyme recognizes a particular amino acid, as well as all the transfer RNAs that

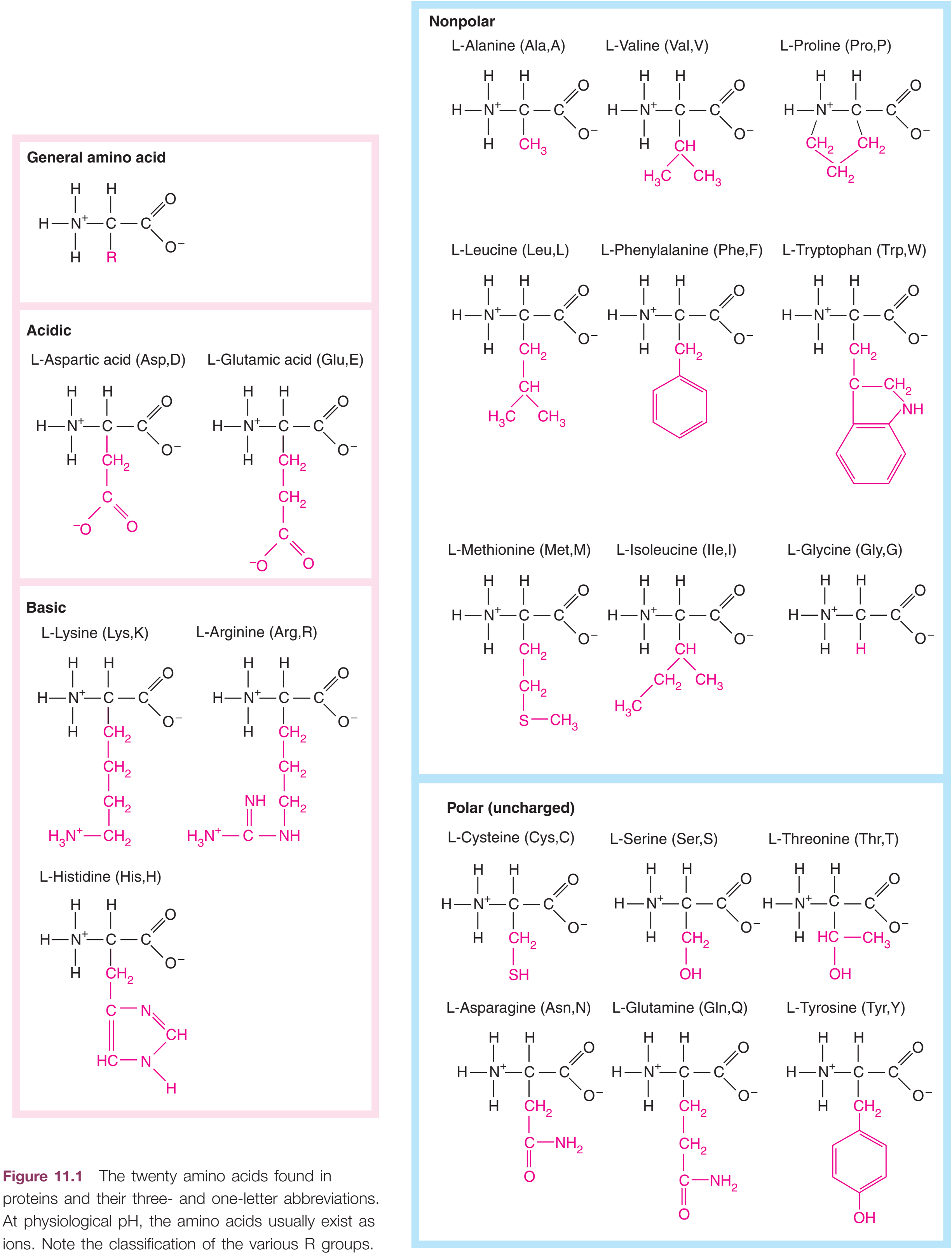

General amino acid
Acidic
L-Aspartic acid (Asp,D)
L-Glutamic acid (Glu,E)
Basic
L-Lysine (Lys,K)
L-Arginine (Arg,R)
L-Histidine (His,H)
Nonpolar
L-Alanine (Ala,A)
L-Valine (Val,V)
L-Proline (Pro,P)
L-Leucine (Leu,L)
L-Phenylalanine (Phe,F)
L-Tryptophan (Trp,W)
L-Methionine (Met,M)
L-Isoleucine (Ile,I)
L-Glycine (Gly,G)
Polar (uncharged)
L-Cysteine (Cys,C)
L-Serine (Ser,S)
L-Threonine (Thr,T)
L-Asparagine (Asn,N)
L-Glutamine (Gln,Q)
L-Tyrosine (Tyr,Y)

**Figure 11.1** The twenty amino acids found in proteins and their three- and one-letter abbreviations. At physiological pH, the amino acids usually exist as ions. Note the classification of the various R groups.

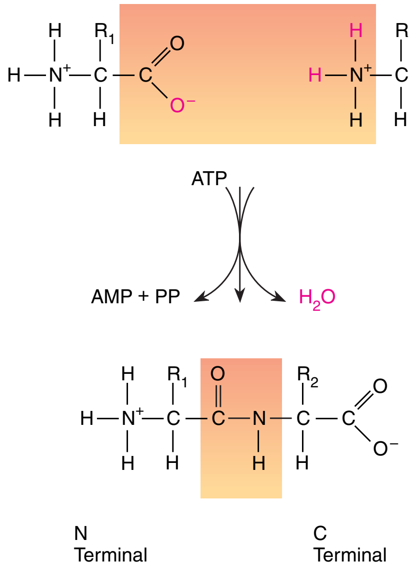


**Figure 11.2** Protein synthesis: formation of a peptide bond between two amino acids. The bond is between the carboxyl group of one amino acid and the amino group of the other.

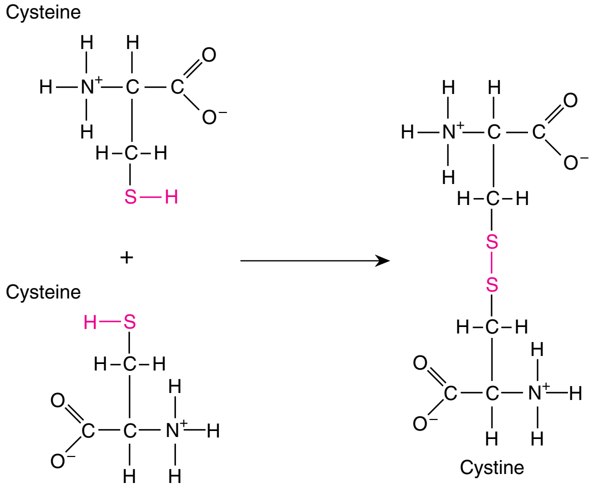


**Figure 11.3** A disulfide bridge can form when two cysteines are brought into apposition. If the two amino acids are in the free form, the new structure is called *cystine.* When the two cysteines are in the same or different polypeptides, the disulfide bridge creates stability.

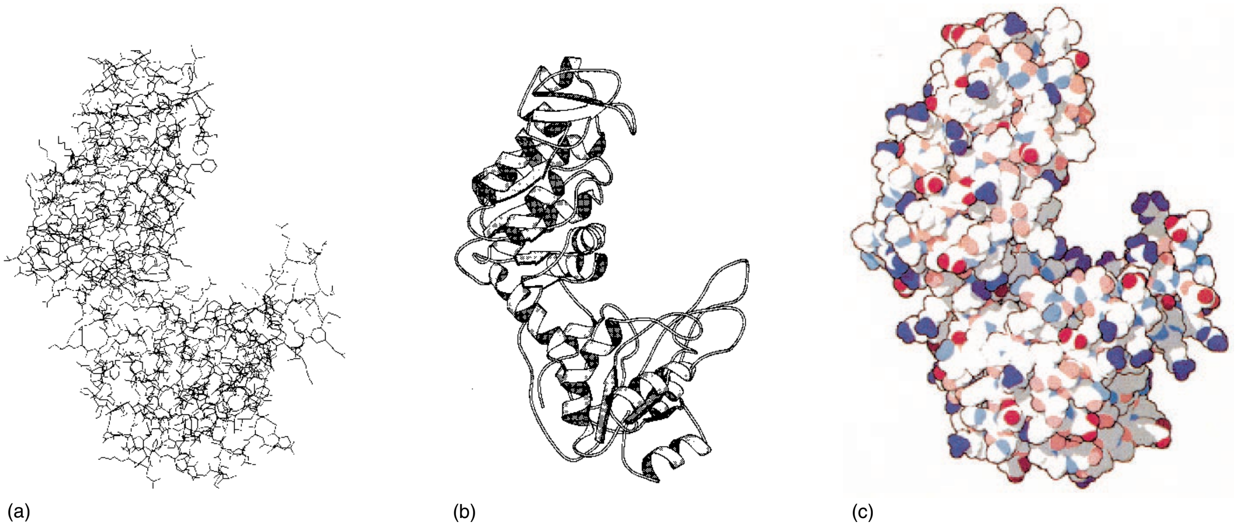


**Figure 11.4** Three different ways of depicting a protein, the enzyme phosphoglycerate kinase. At *left* is a bond diagram; all the lines shown represent bonds between the various atoms of the molecule. In the *middle* is a ribbon diagram that emphasizes the secondary structure of the protein. Shown are alpha helices (spiral ribbons) and beta pleated sheets (flat arrows). Finally, on the *right* is a space-filling diagram that emphasizes the volume the molecule fills. The space-filling diagram is what the molecule would generally look like if it were magnified eight million times. (Images by David S. Goodsell, the Scripps Research Institute.)

## BOX 11.1

## Experimental Methods

### *Amino Acid Sequencing*

Protein-sequencing techniques have been known since 1953, when F. Sanger worked out the complete sequence of the protein hormone insulin. The basic strategy is to purify the protein and then sequence it, beginning at one end. However, since most proteins contain too many amino acids to do this successfully, proteins are first broken into small peptides in several different ways. These peptides are sequenced, and the whole protein sequence can be determined by the overlap pattern of the sequenced subunits.

A protein can be broken into peptide fragments by many different methods, including acid and alkaline hydrolysis. For the most part, proteolytic enzymes (proteases) that hydrolyze the peptides at specific points are used. *Pepsin,* for example, preferentially hydrolyzes peptide bonds involving aromatic amino acids, methionine, and leucine; *chymotrypsin* hydrolyzes peptide bonds involving carboxyl groups of aromatic amino acids; and *trypsin* hydrolyzes bonds involving the carboxyl groups of arginine and lysine.

The proteolytic digest is usually separated into a *peptide map,* or *peptide fingerprint,* by using a two-dimensional combination of paper chromatography, electrophoresis, or column chromatography. In two-dimensional chromatography, a sample is put onto a piece of paper that is then placed in a solvent system. After an allotted time, the paper is dried, turned 90 degrees, and placed in a second solvent system for another allotted time (fig. 1). In each solvent, different peptides travel through the paper at different rates. The spots are then developed using ninhydrin, which reacts with the N-terminal amino acid and produces a colored product when heated.

The spots, which represent small peptides, can be cut out of a second, identical chromatogram that has not been sprayed with ninhydrin. These spots can then be sequenced by, for example, the Edman method, whereby the peptide is sequentially degraded

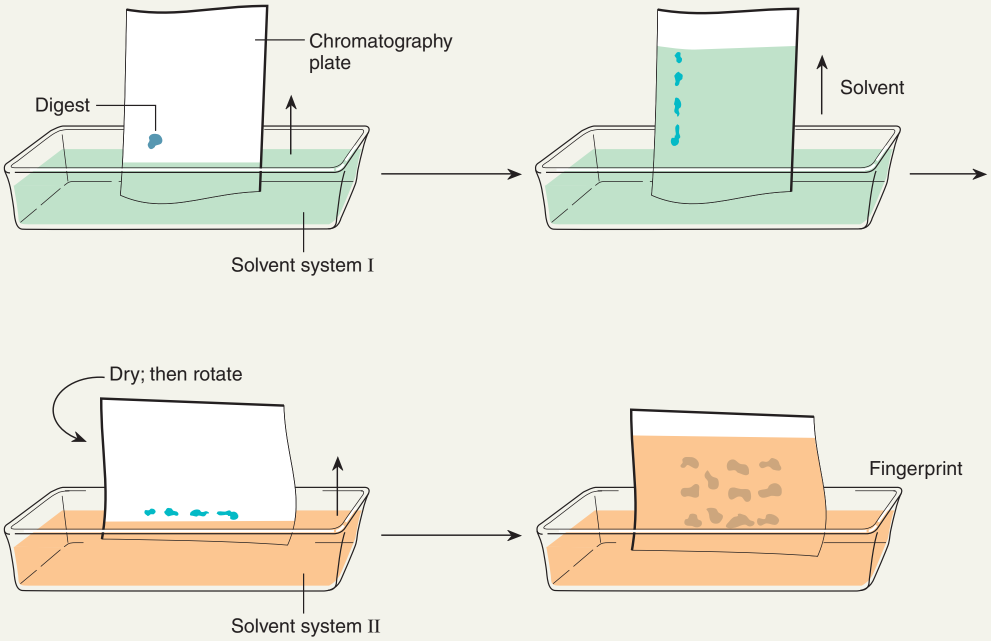


**Figure 1** Two-dimensional paper chromatography of a protease digest. Chromatography is done first in one solvent system. The paper is then dried, rotated, and placed into a second solvent system. The pattern on the resulting plate is called a peptide fingerprint.

from the N-terminal end. Phenylisothiocyanate (PITC) reacts with the amino end of the peptide. When acid is added, the N-terminal amino acid is removed as a PITC derivative and can be identified. The process is then repeated until the whole peptide has been sequenced (fig. 2).

If the fingerprint pattern is worked out for two different digests of the same polypeptide, the unique sequence of the original polypeptide can be determined by overlap. In figure 3, the letters A–J represent the ten amino acids in a polypeptide. A is known to be the first (N-terminal) amino acid since the Edman method sequences peptides from this end. We can thus summarize the methodology as follows:

1. A protein is purified. If it is made up of several subunits, these subunits are separated and purified. (If disulfide bridges exist within a peptide, they must be reduced. The bridges are later determined by digestion, keeping the bridges intact, and then resequencing.)
2. Different proteolytic enzymes are used on separate subsamples so that the protein is broken into different sets of peptide fragments.
3. Two-dimensional chromatography, electrophoresis, or column chromatography can be used to isolate the peptides.
4. The Edman method of sequentially removing amino acids from the N-terminal end is used to sequence each peptide.
5. The amino acid sequence from the N- to C-terminal ends of the protein is deduced from the overlap of sequences in peptide digests generated with different proteolytic enzymes.

Today, a machine known as an amino acid sequencer (*sequenator*) can automatically sequence protein. Taking about two hours per amino acid residue, sequenators can carry out Edman degradation on polypeptides up to about fifty amino acids long.

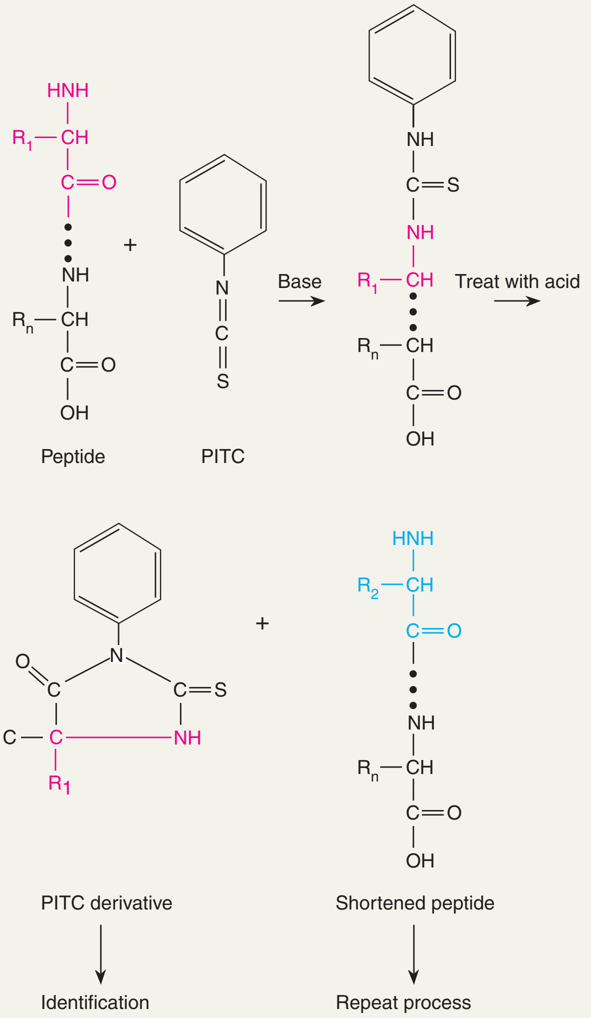


**Figure 2** Isolation of amino acids from a peptide for sequencing purposes. First, the peptide reacts with PITC (phenylisothiocyanate) at the amino end. Acid treatment produces a PITC derivative of the amino-terminal amino acid and a peptide one amino acid shorter than the original. The PITC derivative can be identified. These steps are then repeated, isolating one amino acid at a time.

*continued*

## BOX 11.1 CONTINUED

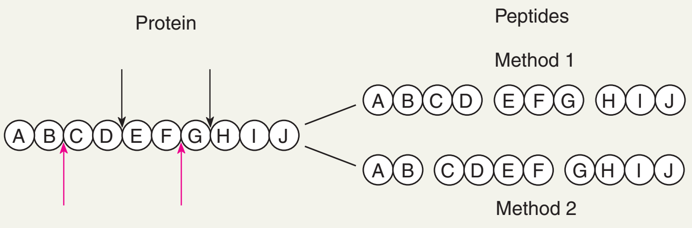


**Figure 3** The overlap of peptides digested two different ways provides the sequence of the original peptide.

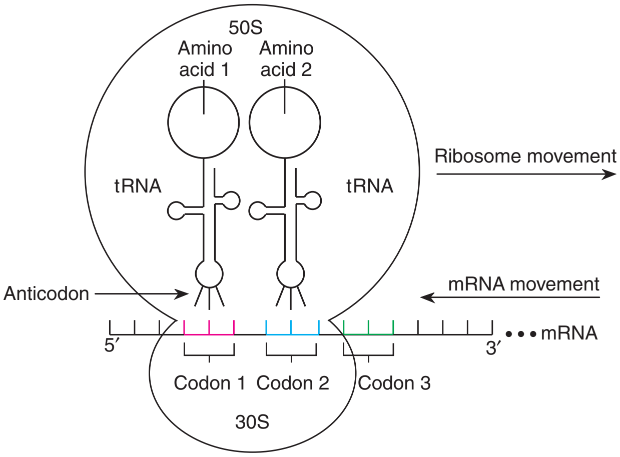


**Figure 11.5** The initiation of the translation process at the ribosome. Note the two charged transfer RNAs and the messenger RNA. They are in position to form the first peptide bond between the two amino acids attached to the transfer RNAs.

code for that amino acid. In eukaryotes, there are separate sets of twenty cytoplasmic and twenty mitochondrial synthetases, all coded in the nucleus.

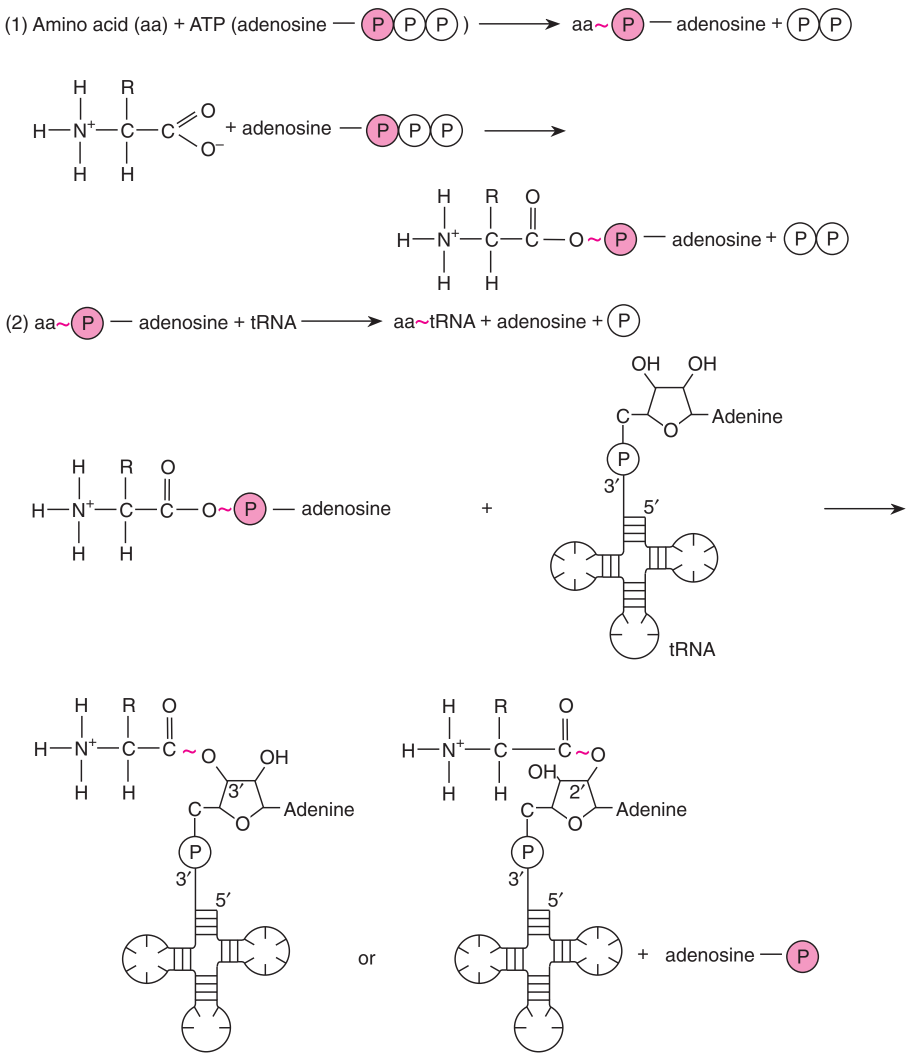
Aminoacyl-tRNA synthetases are a heterogeneous group of enzymes. In *E. coli,* they vary from monomeric proteins (one subunit) to tetrameric proteins, made up of two copies each of two subunits. The enzymes fall into two categories based on sequence similarity, structural features, and whether the amino acid is attached at the 2′-OH (in class I enzymes) or 3′-OH (in class II enzymes) of the 3′-terminal adenosine of the transfer RNA.

To add its appropriate amino acid to the appropriate transfer RNA, a synthetase recognizes many parts of the transfer RNA. This can be shown by experiments that introduce specific changes in transfer RNAs by site-directed mutagenesis (see chapter 13). In seventeen of the twenty *E. coli* synthetases, recognition involves part of the anticodon itself. This makes sense since the anticodon is the defining element of a transfer RNA in protein synthesis.

A synthetase can initially make errors and attach the "wrong" amino acid to a tRNA. For example, isoleucyl-tRNA synthetase will attach valine about once in 225 times. This type of error occurs because a similar, but smaller, amino acid can sometimes occupy the active site of the enzyme (compare isoleucine and valine in fig. 11.1). However, because of a proofreading step, only 1 in 270 to 1 in 800 of the errors are released intact from the enzyme. The amino acids on the rest of the incorrectly charged transfer RNAs are hydrolyzed before the transfer RNAs are released. The overall error rate is the product of the two steps; this means only about one incorrectly charged transfer RNA occurs per 60,000 to 80,000 formed.

In several cases, the number of amino acyl-tRNA synthetases in a particular organism is below twenty. For example, in some archaea, there is no cysteinyl-tRNA synthetase. However, the prolyl-tRNA synthetase activates the tRNAs for both cysteine and proline with their appropriate amino acids. Similarly, in some eubacteria, there is no glutaminyl-tRNA synthetase; the glutaminyl tRNA is charged with glutamic acid, rather than glutamine. An amido transferase enzyme then converts the glutamic acid to glutamine (see fig. 11.1).

There are sixty-four possible codons in the genetic code (four nucleotide bases in groups of three $= 4 \times 4 \times 4 = 64$). Three of these codons are used to terminate translation. Thus, sixty-one transfer RNAs are needed because there are sixty-one different nonterminator codons. About fifty transfer RNAs are known in *E. coli.* The number fifty can be explained by the wobble phenomenon, which occurs in the third position of the codon. We examine this phenomenon in the section on the genetic code. The transfer RNAs for each amino acid are designated by the convention $\text{tRNA}^{\text{Leu}}$ (for leucine), $\text{tRNA}^{\text{His}}$ (for histidine), and so on.


(1) Amino acid (aa) + ATP (adenosine — P P P ) → aa~P — adenosine + P P
H R O O⁻ + adenosine — P P P →
H R O H—N⁺—C—C—O~P — adenosine + P P
(2) aa~P — adenosine + tRNA → aa~tRNA + adenosine + P
OH OH C Adenine O P 3′ 5′
tRNA
OH 3′ Adenine
or
OH 2′ Adenine
+ adenosine — P

Charged tRNA

**Figure 11.6** It takes a two-step process to attach a specific amino acid to its transfer RNA by an aminoacyl synthetase. High-energy bonds are indicated by ~. In the first step, an amino acid is attached to AMP with a high-energy bond. In the second step, the high-energy bond is transferred to the tRNA, which is then referred to as "charged." Depending on which class of aminoacyl-tRNA synthetase is involved, the amino acid will be attached to either the 2′ or 3′ carbon of the sugar of the 3′ terminal adenosine.

### *Recognition of the Aminoacyl-tRNA During Protein Synthesis*

Although amino acids enter the protein-synthesizing process attached to transfer RNAs, it was theoretically possible that the ribosome recognized the amino acid itself during translation. A simple experiment was done to determine whether the amino acid or the transfer RNA was recognized.

In 1962, F. Chapeville and colleagues isolated transfer RNA with cysteine attached. They chemically converted the cysteine to alanine by using Raney nickel, a catalytic form of nickel that removes the SH group of cysteine (fig. 11.7). When these transfer RNAs were used in protein synthesis, alanine was incorporated where cysteine should have been, demonstrating that the transfer RNA, not the amino acid, was recognized during protein synthesis. The synthetase puts a specific amino acid on a specific transfer RNA; then, during protein synthesis, the anticodon on the transfer RNA—not the amino acid itself—determines which amino acid is incorporated.

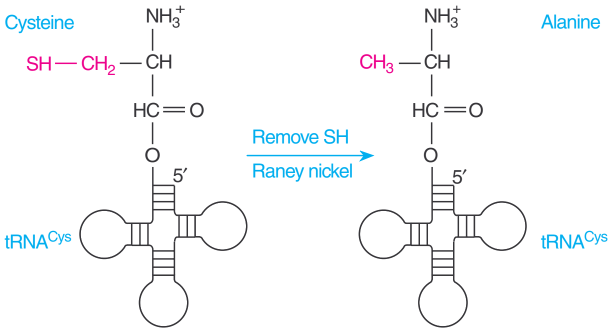


**Figure 11.7** Cysteine-$\text{tRNA}^{\text{Cys}}$ treated with Raney nickel becomes alanine-$\text{tRNA}^{\text{Cys}}$ by the removal of the SH group of cysteine. During protein synthesis, alanine is incorporated in place of cysteine in proteins, indicating that the specificity of amino acid incorporation into proteins resides with the tRNA.

## *Initiation Complex*

Translation can be divided into three stages: initiation, elongation, and termination. Elongation is the repetitive process of adding amino acids to a growing peptide chain. However, added complexity enters the picture in the initiation and termination of protein synthesis.

It is especially important that the translation process start precisely. Remember that the genetic code is translated in groups of three nucleotides (codons). If the reading of the messenger RNA begins one base too early or too late, the reading frame is shifted so that an entirely different set of codons is read (fig. 11.8). The protein produced, if any, will probably bear no structural or functional resemblance to the protein the gene is coded for.

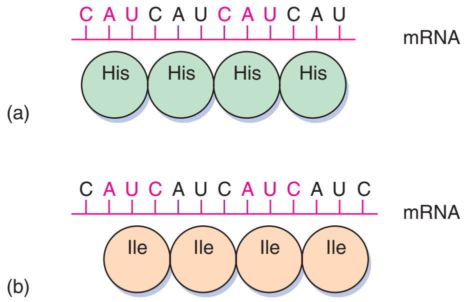


**Figure 11.8** (*a*) In the normal reading of the messenger RNA, these codons are read as repeats of CAU, coding for histidine. (*b*) A shift in the reading frame of the messenger RNA causes the codons to be read as AUC repeats coding for isoleucine.

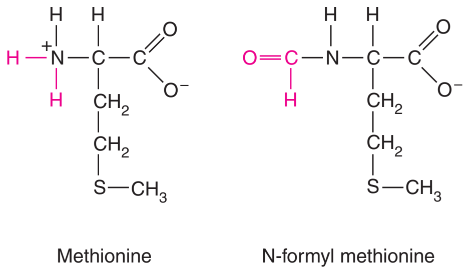


**Figure 11.9** The structures of the amino acids methionine and N-formyl methionine.

### *Role of N-Formyl Methionine*

The synthesis of every protein in *Escherichia coli* begins with the modified amino acid N-formyl methionine (fig. 11.9). However, none of the completed proteins in *E. coli* contains N-formyl methionine. Many of these proteins do not even have methionine as their first amino acid. Obviously, before a protein becomes functional, the initial amino acid is modified or removed. In eukaryotes the initial amino acid, also methionine, does not have an N-formyl group.

Methionine, with a codon of 5′-AUG-3′, known as the **initiation codon,** has two transfer RNAs with the same complementary anticodon (3′-UAC-5′) but with different structures (fig. 11.10). One of these transfer RNAs ($\text{tRNA}_{\text{f}}^{\text{Met}}$) serves as a part of the initiation complex. Before the initiation of translation, this transfer RNA will have its methionine chemically modified to N-formyl methionine (fMet). The other transfer RNA will not have its methionine modified ($\text{tRNA}_{\text{m}}^{\text{Met}}$). The translation machinery will use it to insert methionine into proteins, where called for, in all but the first position. The cell thus has a mechanism to make use of methionine in the normal way as well as to use a modified form of it to initiate protein synthesis. Because of the structure of the prokaryotic initiation transfer RNA, it can recognize AUG, GUG, and, rarely, UUG as initiation codons. In eukaryotes, CUG as well as AUG can serve as an initiation codon. Since the initiation methionine is not formylated in eukaryotes, the eukaryotic transfer RNA is designated $\text{tRNA}_{\text{i}}^{\text{Met}}$; there is a separate internal methionine transfer RNA, termed $\text{tRNAm}_{\text{m}}^{\text{Met}}$, in eukaryotes, as in prokaryotes.

### *Translation Initiation*

The subunits of the ribosome (30S and 50S) usually dissociate from each other when not involved in translation. To begin translation, an **initiation complex** forms, consisting of the following components in prokaryotes: the 30S subunit of the ribosome, a messenger RNA, the charged N-formyl methionine tRNA (fMET-$\text{tRNA}_{\text{f}}^{\text{Met}}$), and

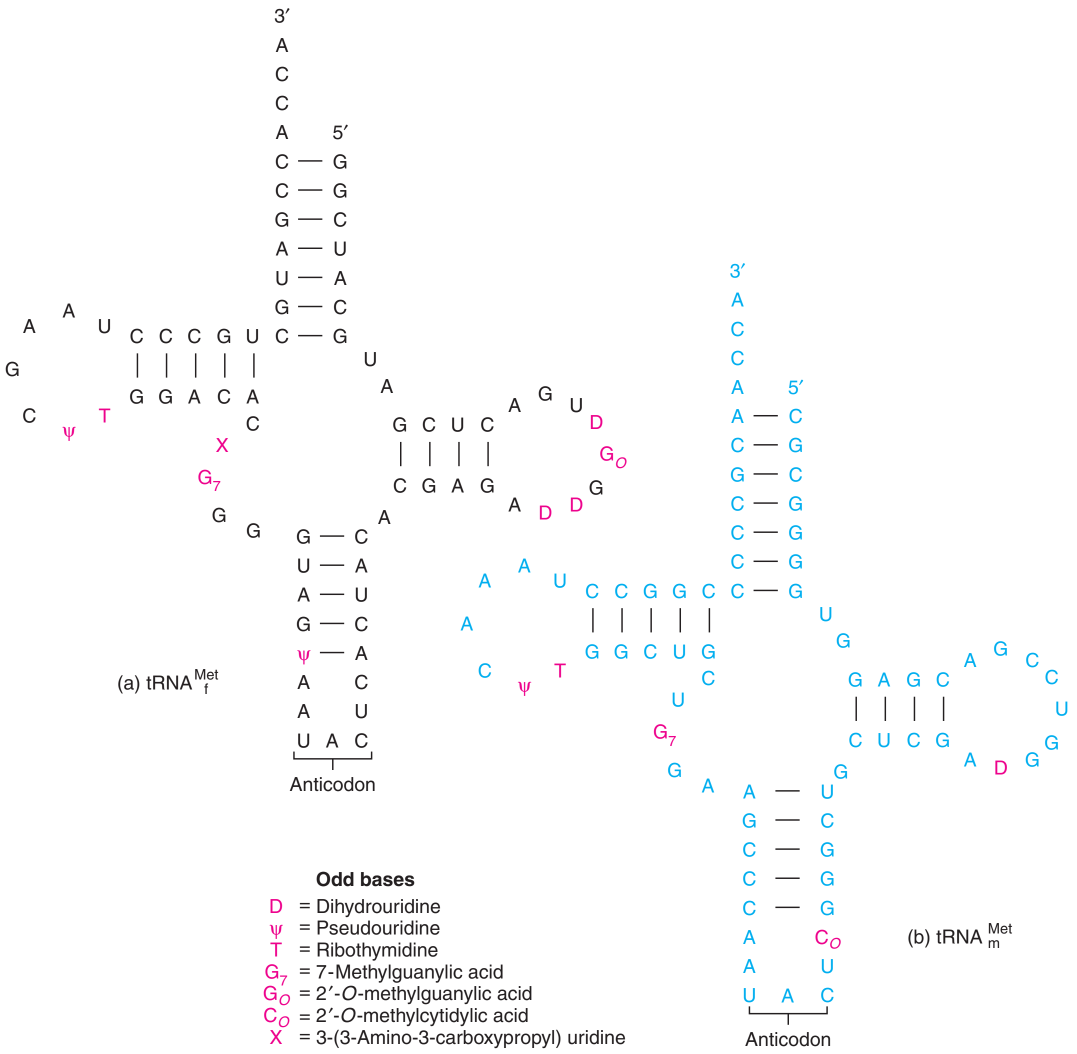

3′
5′
Anticodon
(a) tRNA$_f^{Met}$
3′
5′
Anticodon
(b) tRNA$_m^{Met}$
Odd bases
D = Dihydrouridine
ψ = Pseudouridine
T = Ribothymidine
G$_7$ = 7-Methylguanylic acid
G$_O$ = 2′-*O*-methylguanylic acid
C$_O$ = 2′-*O*-methylcytidylic acid
X = 3-(3-Amino-3-carboxypropyl) uridine

**Figure 11.10** The two tRNAs for methionine in *E. coli.* (*a*) The initiator tRNA. (*b*) The interior tRNA.

three **initiation factors (IF1, IF2, IF3).** Initiation factors (as well as elongation and termination factors) are proteins loosely associated with the ribosome. They were discovered when ribosomes were isolated and then washed, losing the ability to perform protein synthesis.

The components that form the initiation complex interact in a series of steps. It is known that IF3 binds to the 30S ribosomal subunit, allowing the 30S subunit to bind to messenger RNA (fig. 11.11, *step 1*). Meanwhile, a complex forms with IF2, the charged N-formyl methionine tRNA (fMET-tRNA$_f^{Met}$) and GTP (guanosine triphosphate; fig. 11.11, *step 2*). It is IF2 that brings the initiator transfer RNA to the ribosome. IF2 binds only to the charged initiator transfer RNA, and, without IF2, the initiator transfer RNA cannot bind to the ribosome. The final step in initiation-complex formation is bringing together the first two components (fig. 11.11, *step 3*).

The hydrolysis of GTP to GDP + $P_i$ (inorganic phosphate, $PO_4^{-3}$—see fig. 9.8) produces conformational changes; these changes allow the initiation complex to join the 50S ribosomal subunit to form the complete ribosome and then allow the initiation factors and GDP to be released. Frequently, the hydrolysis of a nucleoside triphosphate (e.g., ATP, GTP) in a cell occurs to release the energy in the phosphate bonds for use in a metabolic process. However, in the process of translation, the hydrolysis apparently changes the shape of the GTP so that it and the initiation factors can be released from the ribosome after the 70S particle has been formed. Thus, hydrolysis of GTP in translation is for conformational

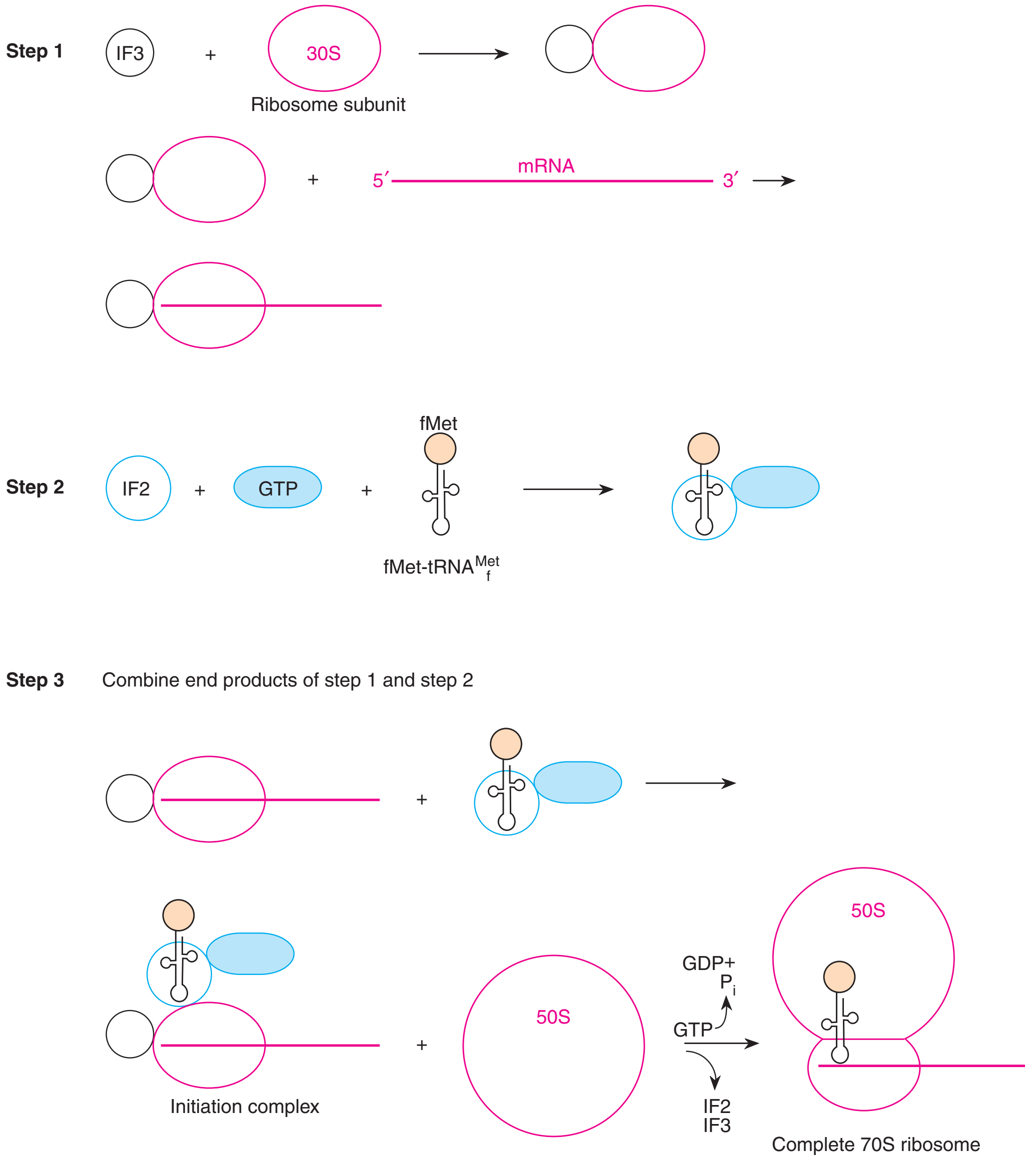

Step 1
IF3
+
30S
Ribosome subunit
+
5′
mRNA
3′
Step 2
IF2
+
GTP
+
fMet
fMet-tRNA$^{Met}_{f}$
Step 3
Combine end products of step 1 and step 2
+
+
50S
GDP+
P$_i$
GTP
IF2
IF3
50S
Initiation complex
Complete 70S ribosome

**Figure 11.11** The prokaryotic 70S ribosome forms in a three-step process. In the first step, the 30S ribosomal subunit and the mRNA combine. In the second step, the initiator tRNA combines with IF2. In the final step, the components from steps 1 and 2 combine to form the initiation complex, followed by the formation of the 70S ribosome.

change rather than covalent bond formation. IF1 helps the other two initiation factors bind to the 30S ribosomal subunit or stabilizes the 30S initiation complex.

The process in eukaryotes is generally similar, but more complex. The eukaryotic initiation factor abbreviations are preceded by an "e" to denote that they are eukaryotic (eIF1, eIF2, etc.). At least eleven initiation factors are involved, including a specific cap-binding protein, eIF4E.

The ribosome apparently recognizes the prokaryotic messenger RNA through complementarity of a region at the 3′ end of the 16S ribosomal RNA and a region slightly upstream from the initiation sequence (AUG) on the messenger RNA. This idea, the **Shine-Dalgarno hypothesis** (fig. 11.12), is named after the people who first suggested it. The sequence (AGGAGGU) of complementarity between the messenger RNA and the 16S ribosomal RNA is referred to as the Shine-Dalgarno sequence. Although

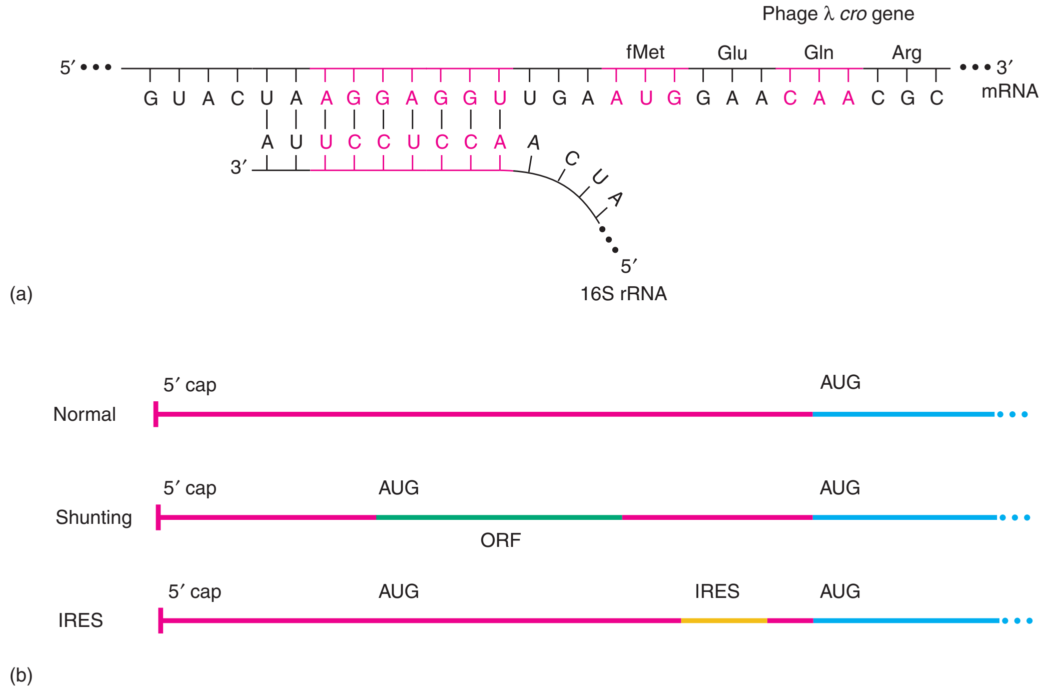


**Figure 11.12** Translation initiation signals. (*a*) The Shine-Dalgarno hypothesis for prokaryotic translation. The Shine-Dalgarno sequence (AGGAGGU) is on the prokaryotic messenger RNA just upstream from the initiation codon AUG. Complementarity exists between this sequence and a complementary sequence (UCCUCCA) on the 3' end of the 16S ribosomal RNA. (*b*) Scanning, shunting, and internal ribosome entry in eukaryotic messenger RNA. The 5' untranslated region of a eukaryotic gene is shown in *red;* the beginning of the gene in *blue.* Normally, in the scanning model, the initiation codon of the gene is the first AUG encountered. In shunting, an open reading frame (ORF, *green*) may or may not be present to provide secondary structure in the messenger RNA to shunt scanning to the main gene. If the open reading frame is translated, reinitiation of translation at the same ribosome may occur at the main gene. Finally, an internal ribosome entry site (IRES, *yellow*) allows translation to begin within the messenger RNA without scanning.

there is a good deal of homology between prokaryotic and eukaryotic small ribosomal RNAs, the Shine-Dalgarno region is absent in eukaryotes. The actual mechanism for recognizing the 5′ end of eukaryotic messenger RNA appears to be based on recognition of the 5′ cap of the messenger RNA by the cap-binding protein with recruitment of other initiation factors and the small subunit of the ribosome. This is followed by the small subunit's movement down the messenger RNA. The ribosome scans the messenger RNA until it recognizes the initiation codon. This model is referred to as the **scanning hypothesis.**

In several known cases in eukaryotes, a process called **shunting** occurs, in which the first AUG does not serve as the initiation codon; rather, scanning begins, but it bypasses a region of the messenger RNA upstream of the initiation codon, called the *leader* or **5′ untranslated region (5′ UTR),** in favor of an AUG further down the messenger RNA. The cause of shunting seems to be secondary structure in the messenger RNA, upstream from the AUG codon that actually serves as the initiation codon. In some cases, very small genes, called **open reading frames (ORFs),** are present in this region of the messenger RNA and play some role in shunting. It seems also that some ORFs are translated, and then the main gene is translated by the same ribosome in a process called **reinitiation** (fig. 11.12*b*). We have seen this in the genes of some plant and animal viruses; it is a topic under current study.

Under some circumstances, eukaryotic ribosomes can initiate protein synthesis within the messenger RNA if that messenger RNA contains a sequence called an **internal ribosome entry site.** These sequences were discovered in the poliovirus RNA and in several cellular messenger RNAs. They are at least four hundred nucleotides long. Thus, although scanning accounts for the initiation of most eukaryotic messenger RNAs at their 5′ ends, some initiation can take place internally in messenger RNAs that have internal ribosome entry sites.

### *Aminoacyl and Peptidyl Sites in the Ribosome*

When the initiator transfer RNA joins the 30S subunit of the prokaryotic ribosome with its messenger RNA attached, it fits into one of three sites in the ribosome. These sites, or cavities in the ribosome, are referred to as the aminoacyl site **(A site),** the peptidyl site **(P site),** and the exit site (**E site,** fig. 11.13) Here, we concentrate on the A and P sites, each of which contains a transfer RNA just before forming a peptide bond: the P site contains the transfer RNA with the growing peptide chain (peptidyl-tRNA); the A site contains a new transfer RNA with its single amino acid (aminoacyl-tRNA). The exit site helps eject depleted transfer RNAs after a peptide bond forms. When the complete 70S ribosome of figure 11.11 has formed, the initiation fMET-$\text{tRNA}_f^{\text{Met}}$ is placed directly into the P site (fig. 11.13), the only charged transfer RNA that can be placed directly there. The association of transfer RNA and ribosome is aided by a G-C base pairing between the 3′-CCA terminus of all transfer RNAs and a guanine in the 23S ribosomal RNA.

## Elongation

### *Positioning a Second Transfer RNA*

The next step in prokaryotic translation is to position the second transfer RNA, which is specified by the codon at the A site. The second transfer RNA is positioned in the A site of the ribosome so it is able to form hydrogen bonds between its anticodon and the second codon on the messenger RNA. This step requires the correct transfer RNA, another GTP, and two proteins called **elongation factors** (**EF-Ts** and **EF-Tu**). EF-Tu, bound to GTP, is required to position a transfer RNA into the A site of the ribosome (fig. 11.14). After the transfer RNA is positioned, the GTP is hydrolyzed to GDP + $P_i$. Upon hydrolysis of the GTP, the EF-Tu/GDP complex is released from the ribosome. EF-Ts is required to regenerate an EF-Tu/GTP complex. EF-Ts displaces the GDP on EF-Tu. Then a new GTP displaces EF-Ts, and now the EF-Tu/GTP complex can bind another transfer RNA. Here again, the hydrolysis of GTP changes the shape of the GTP so that the EF-Tu/GDP complex can depart from the ribosome after the transfer RNA is in place in the A site (fig. 11.15). Figure 11.16 shows the ribosome at the end of this step. EF-Tu does not bind fMet-$\text{tRNA}_f^{\text{Met}}$, so this blocked (formylated) methionine cannot be inserted into a growing peptide chain.

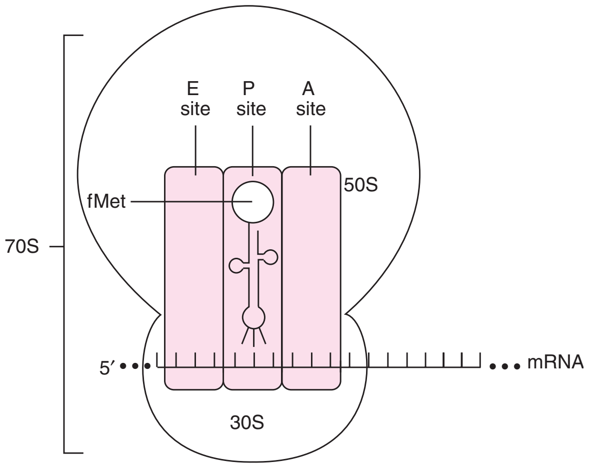


**Figure 11.13** The 70S ribosome contains an A site, a P site, and an E site that can receive tRNAs. The messenger RNA runs through the bottom of the sites.

It takes several milliseconds for the GTP to be hydrolyzed, and another few milliseconds for the EF-Tu/GDP to actually leave the ribosome. During those two intervals of time, the codon-anticodon fit of the transfer RNA is scrutinized. If the correct transfer RNA is in place, a peptide bond forms. If not, the charged transfer RNA is released and a new cycle of EF-Tu/GTP-mediated testing of transfer RNAs begins. The error rate is only about one mistake in ten thousand amino acids incorporated into protein. The speed of amino acid incorporation is about fifteen amino acids per second in prokaryotes and about two to five per second in eukaryotes.

### *Peptide Bond Formation*

The two amino acids on the two transfer RNAs are now in position to form a peptide bond between them; both amino acids are juxtaposed to an enzymatic center, **peptidyl transferase,** in the 50S subunit. This enzymatic center, an integral part of the 50S subunit, was originally believed to be composed of parts of several of the 50S proteins. Now, however, it is believed to have ribozymic activity, enzymatic activity of the ribosomal RNA of the ribosome. The enzymatic activity involves a bond transfer from the carboxyl end of N-formyl methionine to the amino end of the second amino acid (phenylalanine in fig. 11.16). Every subsequent peptide bond is identical, regardless of the amino acids involved. The energy used is contained in the high-energy ester bond between the transfer RNA in the P site and its amino acid (fig. 11.17). Immediately after the formation of the peptide bond, the transfer RNA with the dipeptide is in the A site, and a depleted transfer RNA is in the P site (box 11.2).

### *Translocation*

The next stage in elongation is translocation of the ribosome in relation to the transfer RNAs and the messenger

RNA. Elongation factor EF-G, earlier called translocase, catalyzes the translocation process. The ribosome must be converted from the *pretranslocational state* to the *posttranslocational state* by the action of EF-G, which physically moves the messenger RNA and its associated transfer RNAs (fig. 11.18). This movement is accomplished by the hydrolysis of a GTP to GDP after EF-G enters the ribosome at the A site. After the first posttranslocational state is reached, the depleted transfer RNA in the E site is ejected, leaving the ribosome ready to accept a new charged transfer RNA in the A site. A computer-generated diagram of a ribosome with all three transfer RNA sites occupied is shown in figure 11.18*b*. In eukaryotes, three elongation factors perform the same tasks that EF-Tu, EF-Ts, and EF-G perform in prokaryotes. The factor eEF1α replaces EF-Tu, eEF1βγ replaces EF-Ts, and eEF2 replaces EF-G.

When translocation is complete, the situation is again as diagrammed in figure 11.13, except that instead of fMet-$tRNA_f^{Met}$, the P site contains the second transfer RNA ($tRNA^{Phe}$) with a dipeptide attached to it. The process of elongation is then repeated, with a third transfer RNA coming into the A site. The process repeats from here to the end (fig. 11.19), synthesizing a peptide starting from the amino (N-terminal) end and proceeding to the carboxyl (C-terminal) end. During the repetitive aspect of

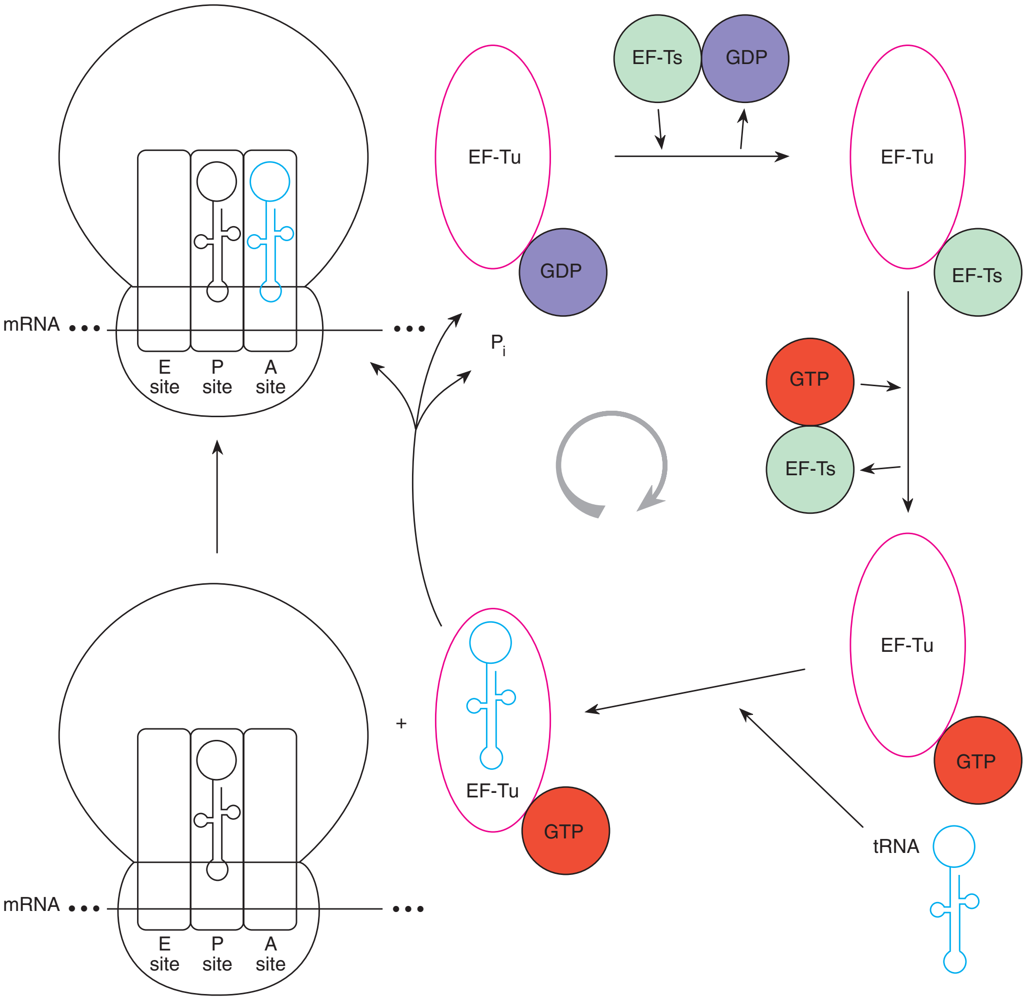


**Figure 11.14** The EF-Ts/EF-Tu cycle. EF-Ts and EF-Tu are required for a transfer RNA to attach to the A site of the ribosome. At *top center,* we have EF-Tu attached to a GDP. The GDP is then displaced by EF-Ts, which in turn is displaced by GTP. A transfer RNA attaches and is brought to the ribosome. If the codon-anticodon fit is correct, the transfer RNA attaches at the A site with the help of the hydrolysis of GTP to GDP + $P_i$, allowing the EF-Tu to release. The EF-Tu is now back where we started. Since EF-Tu has a strong affinity for GDP, the role of EF-Ts is to displace the GDP, and later to be replaced by GTP.

## BOX 11.2

## Biomedical Applications

### *Antibiotics*

Antibiotics, substances living organisms produce that are toxic to other living organisms, are of interest to us for two reasons: They have been extremely important in fighting the diseases that strike human beings and farm animals, and many are useful tools for analyzing protein synthesis. Some antibiotics impede the process of protein synthesis in a variety of ways, often poisoning bacteria selectively; the effectiveness of antibiotics normally derives from the metabolic differences between prokaryotes and eukaryotes. For example, an antibiotic that blocks a 70S bacterial ribosome without affecting an 80S human ribosome could be an excellent antibiotic. About 160 antibiotics are known.

#### PUROMYCIN

Puromycin resembles the 3′ end of an aminoacyl-tRNA (fig. 1). It is bound to the A site of the bacterial ribosome, where peptidyl transferase creates a bond from the nascent peptide attached to the transfer RNA in the P site to puromycin. Elongation can then no longer occur. The peptide chain is released prematurely, and protein synthesis at the ribosome terminates.

Experiments with puromycin helped demonstrate the existence of the A and P sites of the ribosome. It was found that puromycin could not bind to the ribosome if translocation factor EF-G were absent. With EF-G, translocation took place, and puromycin could then bind to the ribosome. Puromycin's ability to bind only after translocation indicates that a second site on the ribosome becomes available after translocation.

#### STREPTOMYCIN, TETRACYCLINE, AND CHLORAMPHENICOL

Streptomycin, which binds to one of the proteins (protein S12) of the 30S subunit of the prokaryotic ribosome, inhibits initiation of protein synthesis. Streptomycin also causes misreading of codons if chain initiation has already begun, presumably by altering the conformation of the ribosome so that transfer RNAs are less firmly bound to it. Bacterial mutants that are streptomycin resistant, as well as mutants that are streptomycin dependent (they cannot survive without the antibiotic), occur. Both types of mutants have altered 30S subunits, specifically changed in protein S12.

Tetracycline blocks protein synthesis by preventing an aminoacyl-tRNA from binding to the A site on the ribosome. Chloramphenicol blocks protein synthesis by binding to the 50S subunit of the prokaryotic ribosome, where it blocks the peptidyl transfer reaction. Chloramphenicol does not affect the eukaryotic ribosome. However, chloramphenicol, as well as several other antibiotics, is used cautiously because the mitochondrial ribosomes within eukaryotic cells are very similar to prokaryotic ribosomes. Some of the antibiotics that affect prokaryotic ribosomes thus also affect mitochondria. As was mentioned, the similarity between bacteria and mitochondria implies that mitochondria have a prokaryotic origin. (Similarities between cyanobacteria and chloroplasts also support the idea that chloroplasts have a prokaryotic origin.)

#### THE TROUBLE WITH ANTIBIOTICS

Over the years, antibiotics have virtually eliminated certain diseases from the industrialized world. They have also made modern surgery possible by preventing most serious infections that tend to follow operations. Antibiotics have been so successful that, in the 1980s, many pharmaceutical companies drastically cut back the development of new antibiotics. However, a disaster was in the making as we overprescribed antibiotics to people and farm animals: bacteria are not prepared to take this onslaught without fighting back.

Mutation takes place all the time at a low but dependable rate. Thus, resistant bacteria are constantly arising from sensitive strains. We can select for penicillin- and streptomycin-resistant strains of bacteria in the laboratory by allowing the antibiotic to act as a selective agent, removing all but the resistant individuals. The same sort of artificial selection that we can apply in the lab applies every time a person or animal takes an antibiotic. We may be at a point now where the ability of bacteria to develop resistance, and to pass that resistance to other strains, has put us on the verge of disaster. The process of evolution works amazingly fast in bacteria because of their ubiquity, large population sizes, and ability to transfer genetic material between individuals. We may shortly find ourselves as we were before World War II, when simple infections in hospitals were often lethal. Right now, only one antibiotic can keep the common—and potentially deadly—infectious bacterium *Staphylococcus* under control: vancomycin. Several types of disease-causing bacteria have already evolved a tolerance to vancomycin.

The answer to this potentially disastrous problem is to develop new antibiotics and reduce the irresponsible use of antibiotics in people and animals. Hopefully, the warning bell has sounded. At least a dozen new antibiotics that show promise are in the early stages of development by pharmaceutical companies.

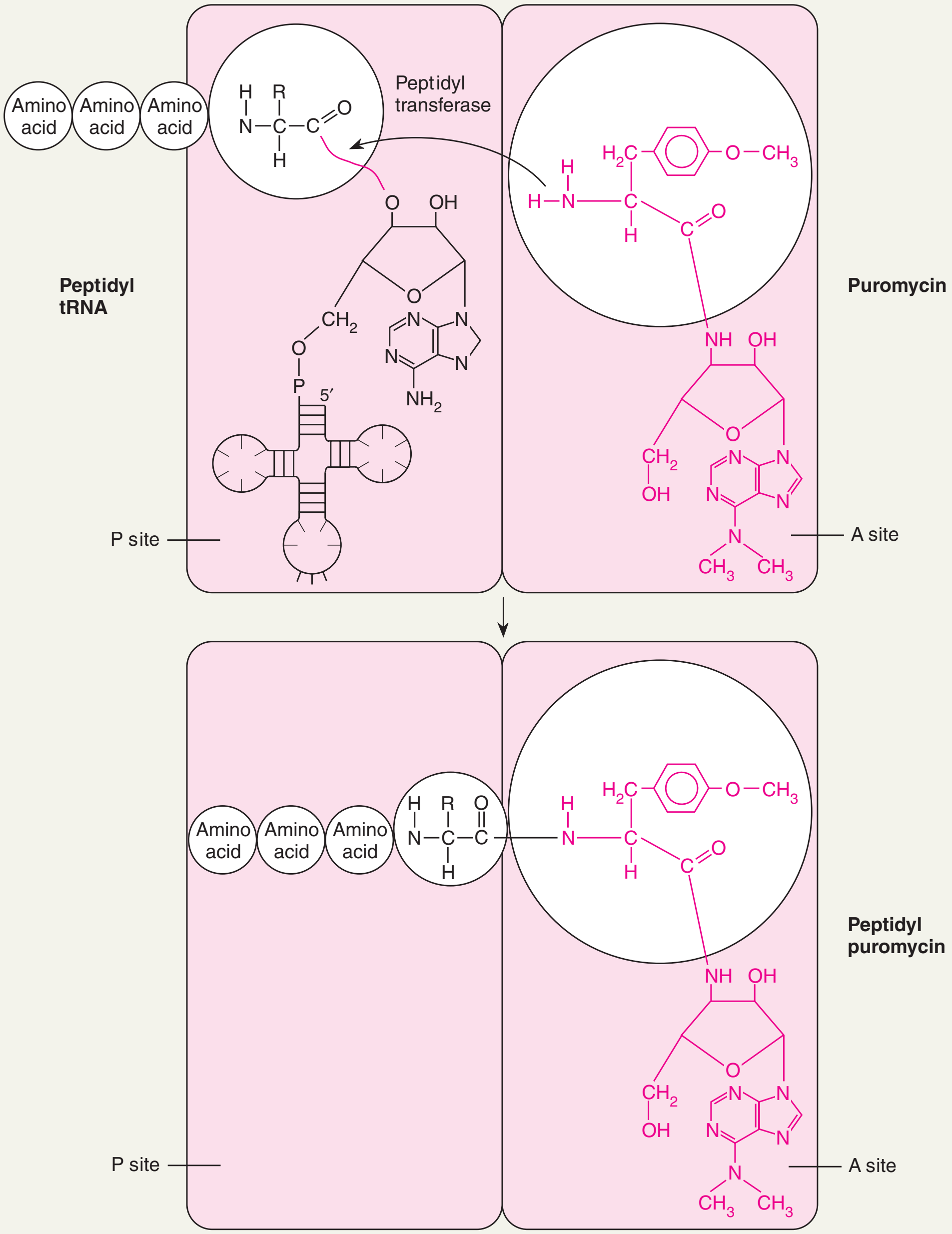

Amino acid
Amino acid
Amino acid
H R
N C C O
H
Peptidyl transferase
O OH
O
CH2
O
P 5′
N N
N N
NH2
Peptidyl tRNA
P site
H
H N C
H2C
O CH3
H
C O
NH OH
O
CH2
OH
N N
N N
N
CH3 CH3
Puromycin
A site
Amino acid
Amino acid
Amino acid
H R O
N C C
H
Peptidyl puromycin
P site
A site

**Figure 1** Puromycin is bound to the A site of the ribosome. A peptide bond then forms. Further elongation is prevented, and the chain is terminated.

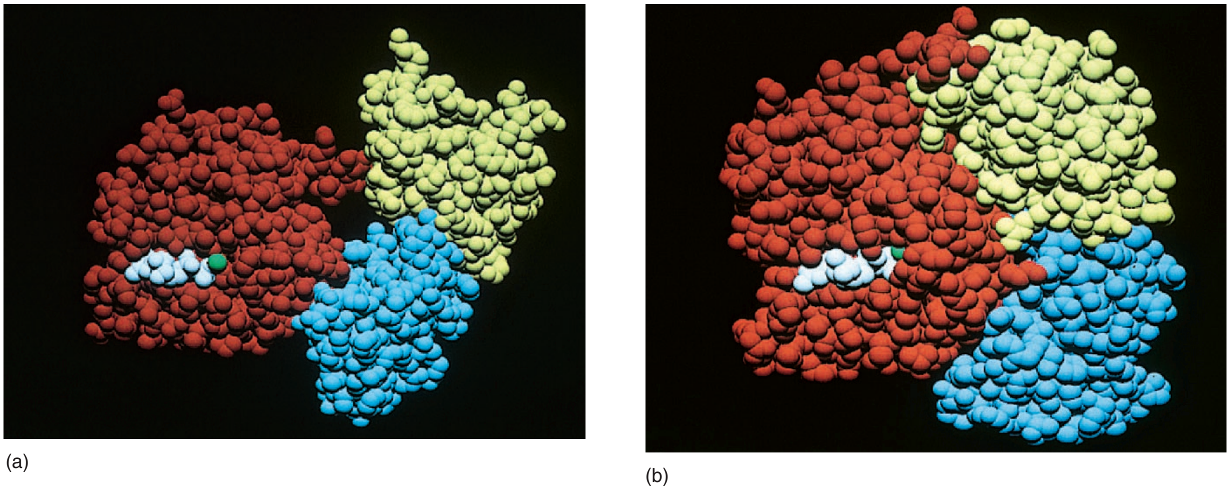


**Figure 11.15** Space-filling model of EF-Tu bound with (*a*) GDP and (*b*) GTP, showing the change in the protein's structure. *Yellow, blue,* and *red* are domains of the protein. The GTP and GDP are in *white,* with a magnesium ion, $Mg^{2+}$, in *green.* When EF-Tu is bound with GDP, there is a visible hole in the molecule. The hole disappears when GTP is bound. The aminoacyl-transfer RNA is believed to bind between the *red* and *yellow* domains. (Courtesy of Rolf Hilgenfeld.)

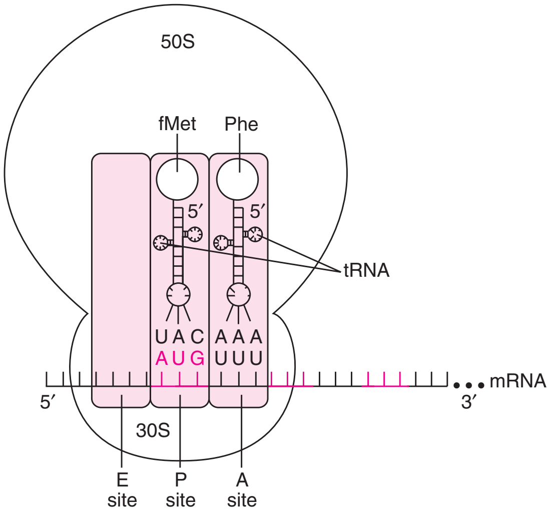


**Figure 11.16** A ribosome with two transfer RNAs attached. In this case, the second codon (UUU) is for the amino acid phenylalanine. The two amino acids are next to each other.

protein synthesis, two GTPs are hydrolyzed per peptide bond: one GTP in the release of EF-Tu from the A site, and one GTP in the translocational process of the ribosome after the peptide bond has formed. In addition, every charged transfer RNA has had an amino acid attached at the expense of the hydrolysis of an ATP to AMP +PP. There is some evidence that the action of EF-Tu hydrolyzes two GTPs.

## Termination

### Nonsense Codons

Termination of protein synthesis in both prokaryotes and eukaryotes occurs when one of three **nonsense codons** appears in the A site of the ribosome. These codons are UAG (sometimes referred to as *amber*), UAA (*ochre*), and UGA (*opal*). ("Amber," or brown stone, is the English translation of the name Bernstein, a graduate student who took part in the discovery of UAG in R. H. Epstein's lab at the California Institute of Technology. "Ochre" and "opal" are tongue-in-cheek extensions of the first label.) In prokaryotes, three proteins called **release factors (RF)** are involved in termination, and a GTP is hydrolyzed to GDP + $P_i$.

When a nonsense codon enters the A site on the ribosome, a release factor recognizes it. RF1 and RF2 are class 1 release factors: They recognize stop codons and then promote hydrolysis of the bond between the terminal amino acid and its tRNA in the P site. Class 2 release factors (RF3) do not recognize stop codons, but they stimulate class 1 release factors to act. RF1 recognizes the stop codons UAA and UAG, and RF2 recognizes UAA and UGA (fig. 11.20). Both do so because they have tripep-

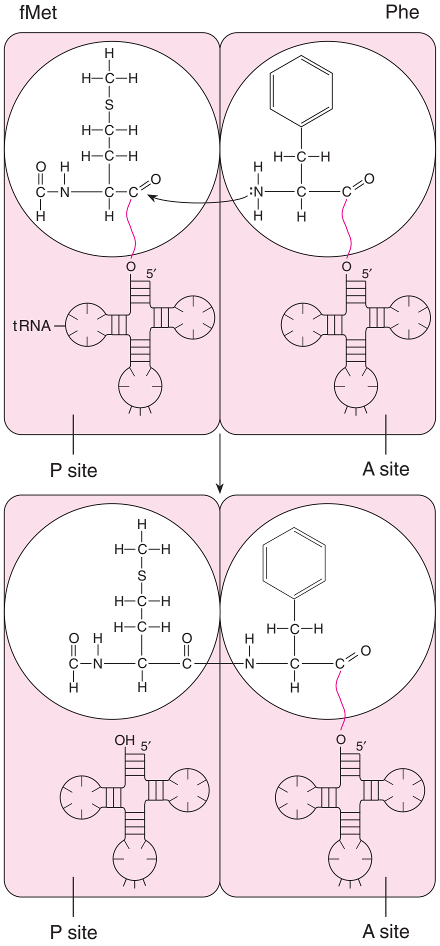


**Figure 11.17** Peptide bond formation on the ribosome between N-formyl methionine and phenylalanine. The bond attaching the carboxyl end of the first amino acid to its tRNA is transferred to the amino end of the second amino acid. The first tRNA is now uncharged, whereas the second tRNA has a dipeptide attached.

tides that mimic anticodons to recognize the stop codons: proline-alanine-threonine in RF1 and serine-proline-phenylalanine in RF2. In this **molecular mimicry,** a protein mimics the shape of a nucleic acid in order to function properly.

The next base in the messenger RNA past the stop codon is usually an adenine, required for efficient termination. After the release factors act, with the hydrolysis of a GTP, the ribosome has completed its task of translating mRNA into a polypeptide. Final release of all factors and dissociation of the two subunits of the ribosome take place with the help of IF3, which rebinds to the 30S subunit, and a **ribosome recycling factor (RRF).** Table 11.1 compares prokaryotic and eukaryotic translation.

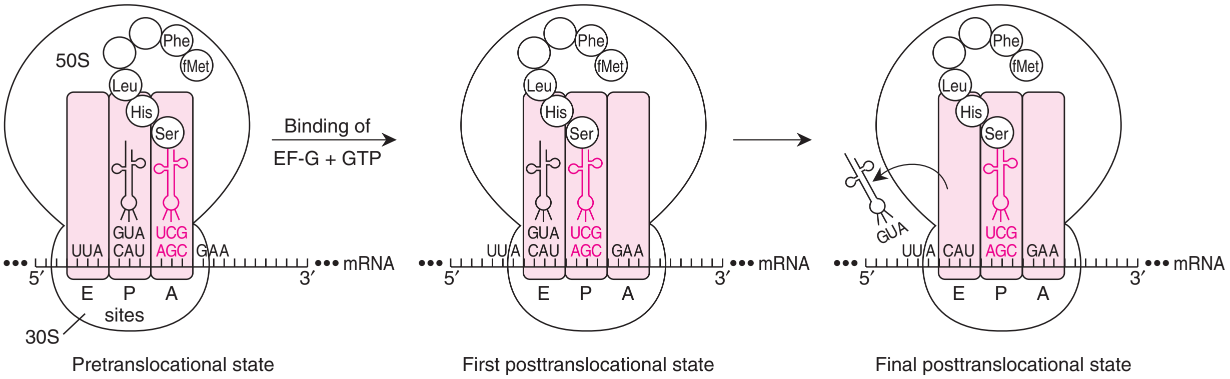
### *Rate and Cost of Translation*

As mentioned, the average speed of protein synthesis is about fifteen peptide bonds per second in prokaryotes. Discounting the time for initiation and termination, an average protein of three hundred amino acids is synthesized in about twenty seconds (the released protein forms its final structure spontaneously or is modified with the aid of other proteins, as we shall see). An equivalent eukaryotic protein takes about 2.5 minutes to be synthesized. The energy cost is at least four high-energy phosphate bonds per peptide bond (two from an ATP during transfer RNA charging, and two from GTP hydrolysis during transfer RNA binding at the A site and translocation), or about twelve hundred high-energy bonds per protein. This cost is very high—about 90% of the energy production of an *E. coli* cell goes into protein synthesis. A high energy cost is presumably the price a living system has to pay for the speed and accuracy of its protein synthesis.

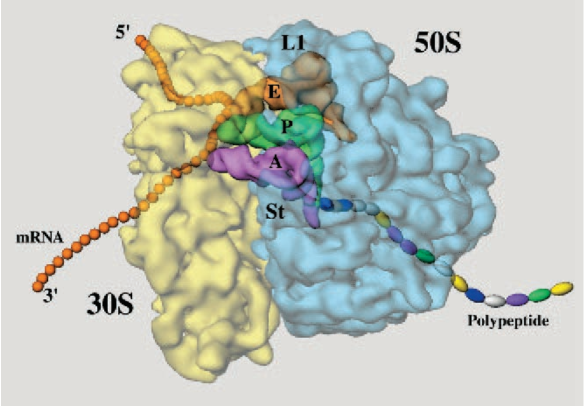
### *Coupling of Transcription and Translation*

In prokaryotes, such as *E. coli,* in which no nuclear envelope exists, translation begins before transcription is completed. Figure 11.21 shows a length of an *E. coli* chromosome. An RNA polymerase is visible on the DNA, transcribing a gene. The messenger RNA, still being synthesized, can be seen extending away from the DNA. Attached to the messenger RNA are about a dozen ribosomes. Since translation starts at the end of the messenger that is synthesized first (5′), an initiation complex can form and translation can begin shortly after transcription begins. As translation proceeds along the messenger, its 5′ end will again become exposed, and a new initiation complex can form. The occurrence of several ribosomes translating the same messenger is referred to as a **polyribosome,** or simply a **polysome** (fig. 11.22).

In prokaryotes, most messenger RNAs contain the information for several genes. These RNAs are said to be **polycistronic** (fig. 11.23). (*Cistron,* another term for gene, is defined in chapter 12.) Each gene on the messenger RNA is translated independently: each has a Shine-Dalgarno sequence for ribosome recognition (see fig. 11.12) and an initiation codon (AUG) for fMet. The ribosome that completes the translation of the first gene may or may not continue to the second gene after dissociation. The translation of any gene follows all the steps we have outlined.


50S
Phe
fMet
Leu
His
Ser
GUA
UCG
UUA
CAU
AGC
GAA
5′
3′
mRNA
E
P
A
sites
30S
Binding of
EF-G + GTP
UU A
Pretranslocational state
First posttranslocational state
Final posttranslocational state

(a)

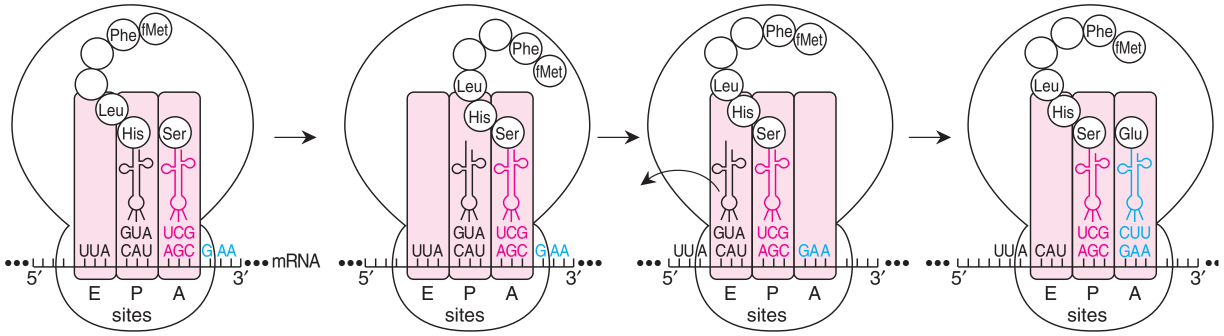

5'
L1
50S
E
P
A
St
mRNA
3'
30S
Polypeptide

(b)

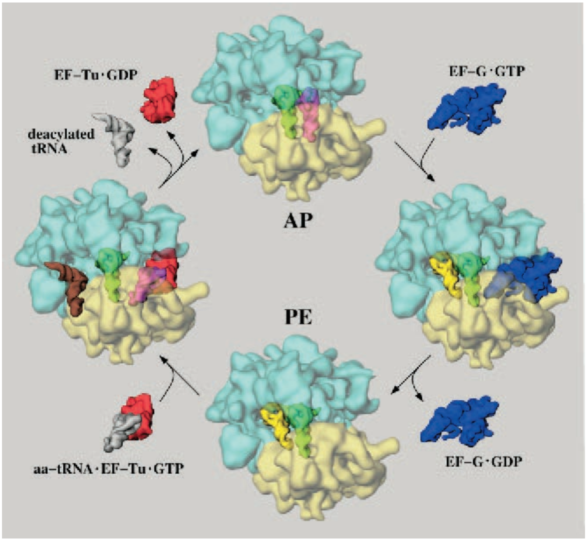
**Figure 11.18** (*a*) EF-G's translocation of the ribosome converts it from a pretranslocational state (P and A sites occupied) to a posttranslocational state (E and P sites occupied). The uncharged transfer RNA in the E site is then ejected. (*b*) A see-through model of the 70S ribosome of *E. coli* with transfer RNAs in the A, P, and E sites. The structure was determined by cryoEM mapping, an electron microscopic technique using rapidly frozen specimens. The position of the messenger RNA is shown, as well as the stalk of the 50S subunit (St) and one of the polypeptides of the large subunit, L1. ([*b*] Courtesy of Joachim Frank, Howard Hughes Medical Institute.)

**Table 11.1** **Some Comparisons Between Prokaryotic and Eukaryotic Translation**

| | Prokaryotes | Eukaryotes |
|---|---|---|
| Initiation codon | AUG, occasionally GUG, UUG | AUG, occasionally GUG,CUG |
| Initiation amino acid | N-formyl methionine | Methionine |
| Initiation tRNA | $tRNA_f^{Met}$ | $tRNA_i^{Met}$ |
| Interior methionine tRNA | $tRNA_m^{Met}$ | $tRNA_m^{Met}$ |
| Initiation factors | IF1,IF2,IF3 | eIF factors |
| Elongation factor | EF-Tu | eEF1α |
| Elongation factor | EF-Ts | eEF1βγ |
| Translocation factor | EF-G | eEF2 |
| Release factors | RF1, RF2, RF3, RRF | eRF1, eRF3 |



(a)



(b)

**Figure 11.19** Cycle of peptide bond formation and translocation on the ribosome. (*a*) After the peptide bond is transferred (fig. 11.17), the ribosome and messenger RNA move over one codon. Now the transfer RNA with the peptide is in the P site, and the A site is again open. In this example, the next transfer RNA that moves into the A site carries glutamic acid. (*b*) Three-dimensional model of the translocation process minus the mRNA and amino acids. The tRNA in the A site is *pale blue,* the tRNA in the P site is *green,* and the tRNA in the E site is *yellow,* then *brown* when ready to leave. Going clockwise from *a,* in which the A and P sites are occupied: EF-G translocates the ribosome after peptide bond formation and then evacuates the A site. Ef-Tu brings a new charged tRNA to the A site while the E site is emptied. ([*b*] Courtesy of Joachim Frank.)

In eukaryotes, however, almost all messenger RNAs contain the information for only one gene **(monocistronic).** Since most ribosomal recognition of eukaryotic genes depends on the 5′ cap, and since each eukaryotic messenger RNA has only one cap, usually only one polypeptide can be translated for any given messenger RNA. Exceptions occur when the messenger RNAs contain internal ribosome entry sites. Although it is certainly not the rule, the translated peptide can be modified or cleaved into smaller functional peptides. For example, in mice, a single messenger RNA codes for a protein that is later cleaved into epidermal growth factor and at least seven other related peptides. In addition, the same sequence can, in some cases, give rise to alternative proteins through alternative start codons, termination read-through, or alternative splicing.

## More on the Ribosome

In the last chapter, we briefly discussed the shape and composition of the ribosomal subunits. All of the protein and RNA components have been isolated. Assembly pathways are known. We know approximately where the messenger RNA, initiation factors, and EF-Tu are located on the 30S subunit during translation (fig. 11.24; cf. fig. 11.18). We also know where peptidyl transferase activity and EF-G reside on the 50S subunit, which has a cleft leading into a tunnel that passes through the structure. At present, it seems that the nascent peptide passes through this

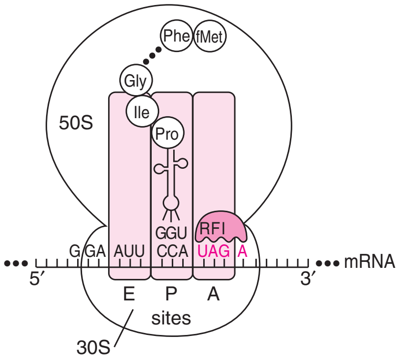

Phe
fMet
Gly
Ile
Pro
50S
GGU
RFI
G GA AUU CCA UAG A
••• mRNA
5′
3′
E P A
sites
30S

**Figure 11.20** Chain termination at the ribosome. One of two release factors recognizes a nonsense codon in the A site. In this case, RF1 recognizes UAG. The complex then falls apart, releasing the peptide.

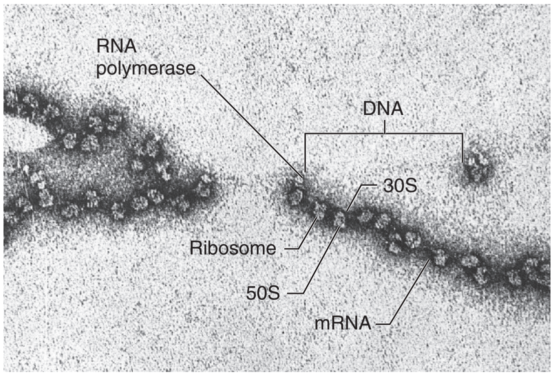

RNA polymerase
DNA
30S
Ribosome
50S
mRNA

**Figure 11.21** A polysome (i.e., multiple ribosomes on the same strand of mRNA). Each ribosome is approximately 250 Å units across. Also visible in this illustration are DNA and RNA polymerase. (Reproduced courtesy of Dr. Barbara Hamkalo, *International Review of Cytology,* (1972) 33:7, fig. 5. Copyright by Academic Press, Inc., Orlando, Florida.)

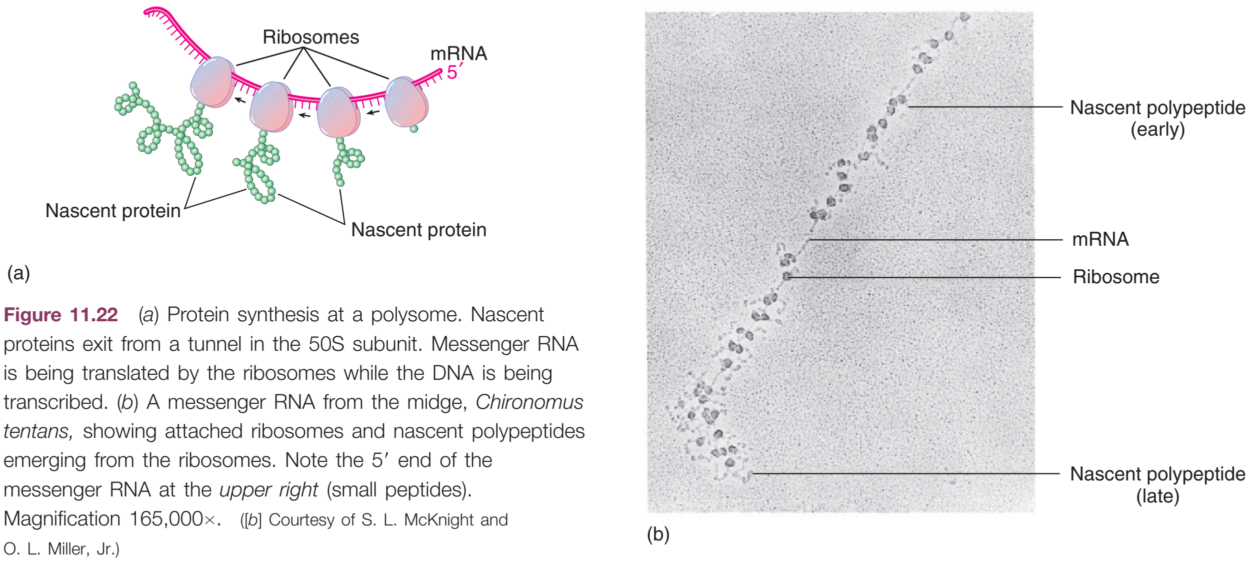

Ribosomes
mRNA
5′
Nascent protein
Nascent protein
(a)
Nascent polypeptide (early)
mRNA
Ribosome
Nascent polypeptide (late)
(b)

**Figure 11.22** (*a*) Protein synthesis at a polysome. Nascent proteins exit from a tunnel in the 50S subunit. Messenger RNA is being translated by the ribosomes while the DNA is being transcribed. (*b*) A messenger RNA from the midge, *Chironomus tentans,* showing attached ribosomes and nascent polypeptides emerging from the ribosomes. Note the 5′ end of the messenger RNA at the *upper right* (small peptides). Magnification 165,000×. ([*b*] Courtesy of S. L. McKnight and O. L. Miller, Jr.)

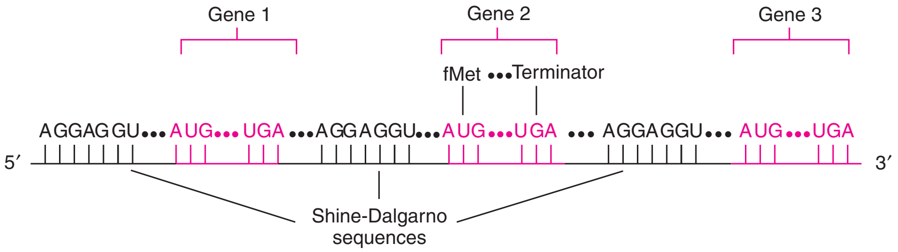

Gene 1
Gene 2
Gene 3
fMet ••• Terminator
AGGAGGU••• AUG••• UGA •••AGGAGGU••• AUG•••UGA ••• AGGAGGU••• AUG•••UGA
5′
3′
Shine-Dalgarno sequences

**Figure 11.23** A prokaryotic polycistronic mRNA. Note the several Shine-Dalgarno sequences for ribosomal attachment and the initiation and termination codons marking each gene.

tunnel, emerging close to a membrane-binding site (fig. 11.24). The tunnel can hold a peptide length of about forty amino acids. Note that although every ribosome has a membrane-binding site, not all active ribosomes are bound to membranes.

Gunter Blobel (1936– ). (Courtesy of Dr. Gunter Blobel, Dept. of Cell Biology, Rockefeller University.)

## The Signal Hypothesis

Ribosomes are either free in the cytoplasm or associated with membranes, depending on the type of protein being synthesized. Membrane-bound ribosomes, indistinguishable from free ribosomes, synthesize proteins that enter membranes. These proteins either become a part of the membrane or, in eukaryotes, either pass into membrane-bound organelles (e.g., the Golgi apparatus, mitochondria, chloroplasts, vacuoles) or are transported outside the cell membrane. The **signal hypothesis** of G. Blobel, a 1999 Nobel laureate, and his colleagues, explains the mechanism for membrane attachment. The mechanism applies to both prokaryotes and eukaryotes. Here, we describe it in mammals.

The signal for membrane insertion is coded into the first one to three dozen amino acids of membrane-bound proteins. This **signal peptide** takes part in a chain of events that leads the ribosome to attach to the membrane and to the insertion of the protein. The first step occurs when the signal peptide becomes accessible outside of the ribosome. A ribonucleoprotein particle called the **signal recognition particle** (SRP), which consists of six different proteins and a 7S RNA about three hundred nucleotides long, recognizes the signal peptide. The complex of signal recognition particle, ribosome, and signal peptide then passes, or diffuses, to a membrane, where the SRP binds to a receptor called a **docking protein** (DP) or signal recognition particle receptor (fig. 11.25). During this time, protein synthesis halts. The ribosome is brought into direct contact with the membrane, and other proteins of the membrane help anchor the ribosome. Protein synthesis then resumes, with the nascent protein usually passing directly into a **translocation channel (translocon).** Once through the membrane, the signal peptide is cleaved from the protein by an enzyme called *signal peptidase.* A striking verification of this hypothesis came about through recombinant DNA techniques (chapter 13). A signal sequence was placed in front of the α-globin gene, whose protein product is normally not transported through a membrane. When this gene was translated, the ribosome became membrane bound, and the protein passed through the membrane.

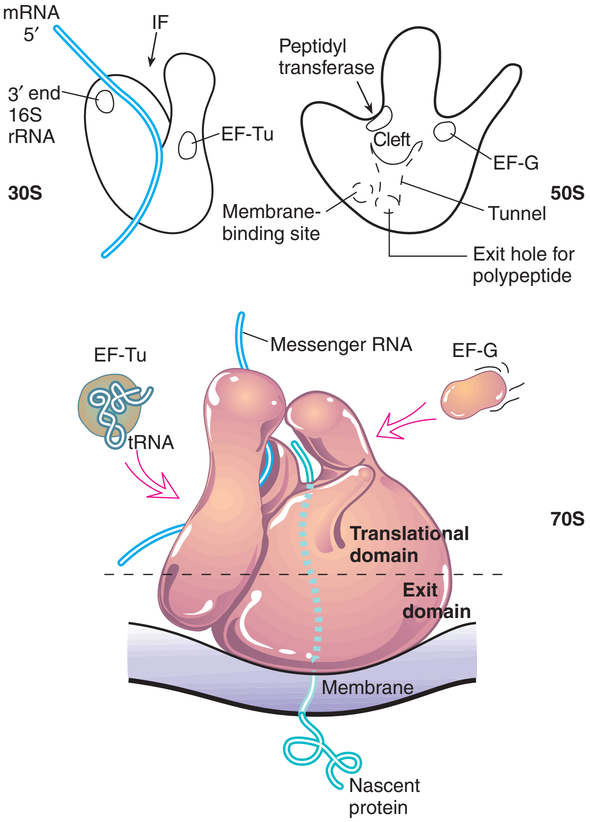


**Figure 11.24** Functional sites on the prokaryotic ribosome. The ribosome is synthesizing a protein involved in membrane passage. Note the position of the messenger RNA on the 30S subunit and the cleft, tunnel, and membrane-binding site on the 50S subunit. (From C. Bernabeu and J. A. Lake, *Proceedings of the The National Academy of Sciences;* 79:3111–15, 1982. Reprinted by permission.)

Since different proteins enter different membrane-bound compartments (e.g., the Golgi apparatus), some mechanism must direct a nascent protein to its proper

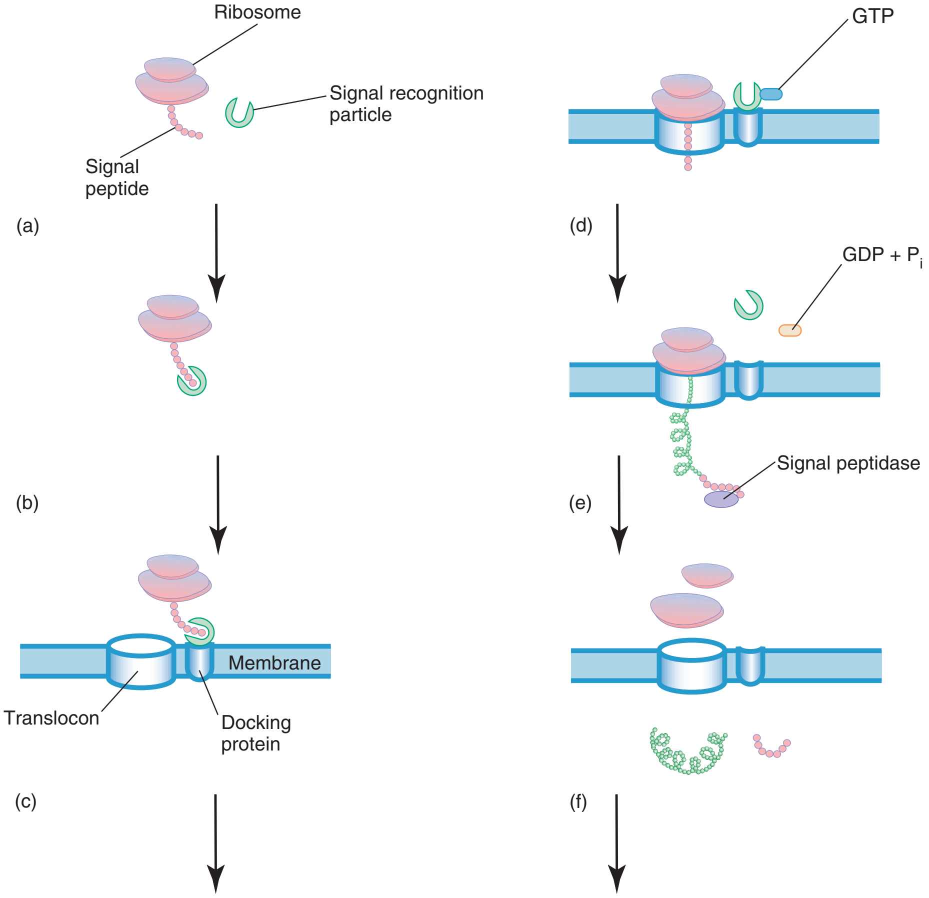


**Figure 11.25** The signal hypothesis. A signal recognition particle recognizes a ribosome with a signal peptide, then draws the ribosome to a docking protein located near a translocon in the membrane. With the addition of GTP, the signal recognition particle releases the signal peptide; hydrolysis of the GTP to GDP + $P_i$ causes the signal recognition particle to leave the docking protein. Peptide synthesis then resumes, with the newly synthesized peptide passing through the translocon in the membrane. A signal peptidase on the other side of the membrane removes the signal peptide. When translation is completed, the ribosome dissociates and drops free of the translocon.

membrane. This specificity seems to depend on the exact signal sequence and membrane-bound glycoproteins called *signal-sequence receptors.* Apparently, after the ribosome binds to the docking protein, the signal peptide interacts with a signal-sequence receptor, which presumably determines whether that protein is specific for that membrane. If it is, the remaining processes continue. If not, the ribosome may be released from the membrane.

The signal peptide does not seem to have a consensus sequence like the transcription or translation recognition boxes. Rather, similarities (at least for the endoplasmic reticulum and bacterial membrane-bound proteins) include a positively charged (basic) amino acid (commonly lysine or arginine) near the beginning (N-terminal end), followed by about a dozen hydrophobic (nonpolar) amino acids, commonly alanine, isoleucine, leucine, phenylalanine, and valine (table 11.2).

### Table 11.2 The Signal Peptide of the Bovine Prolactin Protein*

$NH_2$ − Met Asp Ser Lys Gly Ser Ser Gln Lys Gly Ser Arg Leu
Leu Leu Leu Leu Val Val Ser Asn Leu Leu Leu Cys Gln Gly Val
Val Ser | Thr Pro Val...Asn Asn Cys − COOH

*Source:* From Sasavage et al., *Journal of Biological Chemistry,* 257:678–81, 1982. Reprinted with permission.

* The vertical line separates the signal peptide from the rest of the protein, which consists of 199 residues.

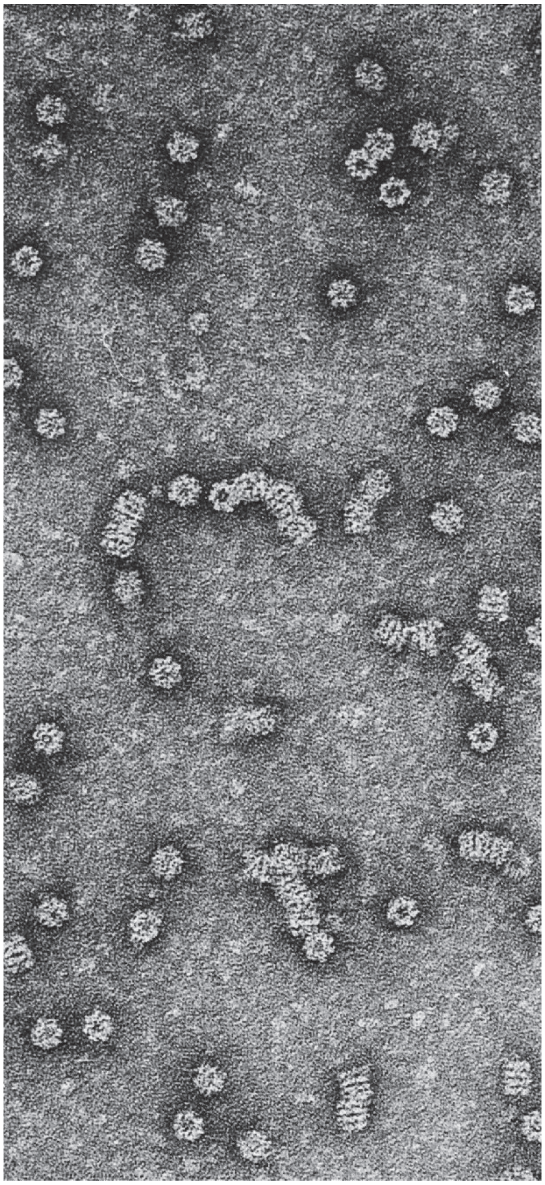


**Figure 11.26** Electron micrograph of a chaperone protein (GroEL) from *E. coli.* Note the hollow, barrel shape of the protein. (Courtesy of Dr. R. W. Hendrix.)

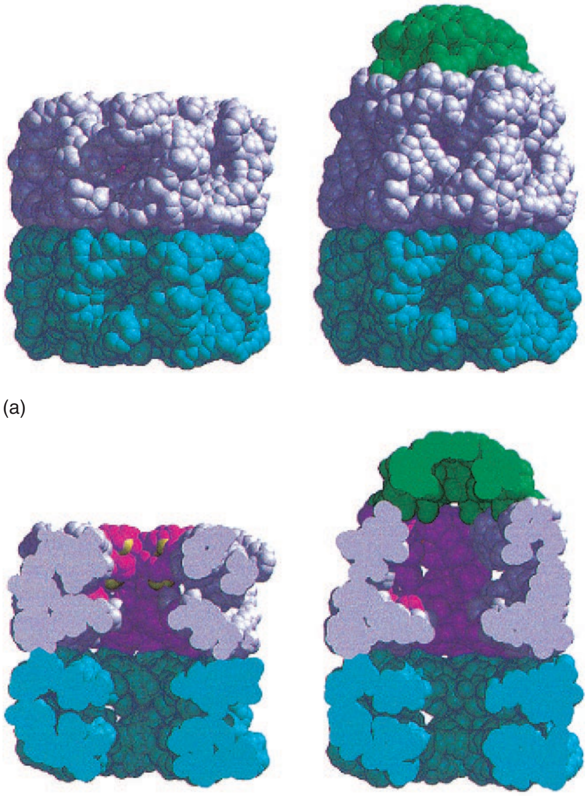
The mitochondrion, which needs to import upwards of one thousand proteins through both inner and outer membranes, poses a specific problem. Recent research has revealed a family of translocation proteins (called Tom proteins) in the outer membrane and a different set of translocation proteins (called Tim proteins) in the inner membrane. These proteins control the passage of proteins synthesized in the cytoplasm into the mitochondrion.

## *The Protein-Folding Problem*

Since biochemist Christian Anfinsen won a 1972 Nobel Prize for showing that the enzyme ribonuclease refolds to its original shape after denaturation in vitro, scientists have believed that the final protein shape (secondary and tertiary structure) forms spontaneously. Recently it has been shown, however, that many proteins do not normally form their final active shape in vivo without the help of proteins called **chaperones** or **molecular chaperones.** The chaperones do not provide the three-dimensional structure of the proteins they help, but rather bind to a protein in the early stages of folding to prevent unproductive folding or to allow denatured proteins to refold correctly. Like human chaperones, they prevent or undo "incorrect interactions," according to J. Ellis. That is, many proteins have a large number of different structures they could fold into. Many of these structures would have no enzymatic activity or would form functionless aggregates with other proteins. Molecular chaperones allow proteins to fold into a thermodynamically stable and functional configuration. Each cycle of refolding requires ATP energy.

A well-studied class of chaperones is known as the *chaperonins,* or Hsp60 proteins, because they are heat shock proteins about 60 kilodaltons (60,000 daltons) in size. They occur in bacteria, chloroplasts, and mitochondria. One of the best studied of these chaperonins is the protein GroE of *E. coli.* This protein in its active form is composed of two components, GroEL and GroES. GroEL (Hsp60) is made up of two disks, each composed of seven copies of a polypeptide. GroES (Hsp10) is a smaller component composed of seven copies of a small subunit. GroEL forms a barrel in which protein folding takes place (fig. 11.26). The barrel is shaped in such a way that entering proteins of a certain size make contact at interior points in either the upper or lower ring of GroEL (upper ring shown in fig. 11.27). The attachment of GroES, the cap, causes the ring to open outward at the top, stretching the protein inside. This stretching takes energy from the hydrolysis of ATP molecules located inside the rings. When GroES dissociates, the protein can fold into a new, more functional, configuration. If it doesn't, the cycle repeats.

There are several classes of molecular chaperones, proteins of different sizes and shapes that recognize different groups of proteins or protein conformations. GroEL recognizes about 300 different proteins, small enough to fit into the barrel (20–60 kilodaltons) and having hydrophobic surfaces. These include many proteins in the transcription and translation machinery of the cell. Hsp90, another heat shock protein, recognizes proteins involved in signal transduction, discussed in chapter 16. Hsp70 recognizes hydrophobic regions in polypeptide side chains, many of which extend across membranes.



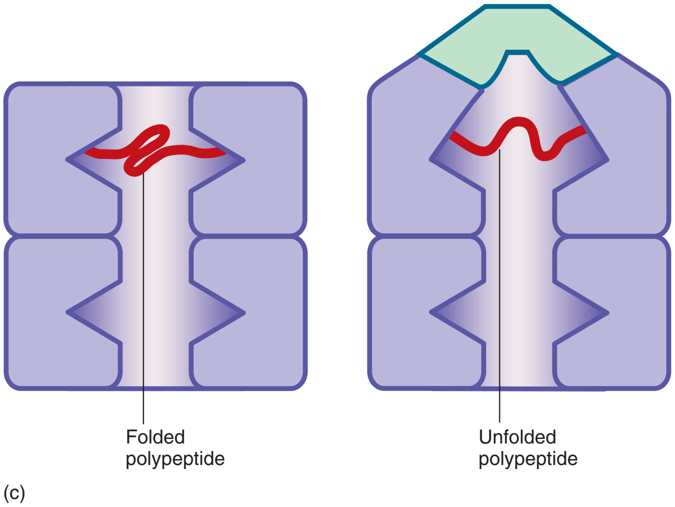


**Figure 11.27** The change in structure of GroEL with GroES attached explains how the chaperonin can unfold a partially folded polypeptide to allow it to refold in a different way. (*a*) A space-filling model of GroEL is shown without (*left*) and with (*right*) GroES. GroEL's rings are *blue* and *magenta,* and GroES is *green.* (*b*) The same structures are seen in a cutaway view. (*c*) This diagram shows how the attachment of GroES causes the top part of the top ring of GroEL to pull apart an improperly folded polypeptide. ([*a* & *b*]: Reprinted from Bernd Bakau and Arthur L. Horwich, "The Hsp70 and Hsp60 Chaperone Machines" in *Cell*, vol. 92, 351–366. Copyright 1988, with permission from Elsevier Science.)

## THE GENETIC CODE

Researchers in the mid-1950s assumed that the genetic code consisted of simple sequences of nucleotides specifying particular amino acids. They sought answers to questions such as: Is the code overlapping? Are there nucleotides between code words (punctuation)? How many letters make up a code word (codon)? Logic, along with genetic experiments, supplied some of the answers, but only with the rapidly improving techniques of biochemistry did they eventually decode the genetic language.

### Triplet Nature of the Code

Several lines of evidence seemed to indicate that the nature of the code was triplet (three bases in messenger RNA specifying one amino acid). If codons contained only one base, they would only be able to specify four amino acids since there are only four different bases in DNA (or messenger RNA). A couplet code would have $4 \times 4 = 16$ two-base words, or codons, which is still not enough to specify uniquely twenty different amino acids. A triplet code would allow for $4 \times 4 \times 4 = 64$ codons, which are more than enough to specify twenty amino acids.

#### Evidence for the Triplet Nature of the Code

The experimental manipulation of mutant genes, primarily by Francis Crick and his colleagues, reinforced the triplet code concept. In these experiments, a chemical mutagen, the acridine dye proflavin, was used to cause inactivation of the rapid lysis (*rIIB*) gene of the bacteriophage T4. Proflavin inactivates the gene by either adding or deleting a nucleotide from the DNA (see chapter 12). The *rII* gene controls the plaque morphology of this bacteriophage growing on *E. coli* cells. Rapid-lysis mutants produce large plaques; the wild-type form of the gene, $rII^+$, results in normal plaque morphology.

Figure 11.28 shows the consequences of adding or deleting a nucleotide. From the point of addition or deletion onward, a **frameshift** causes codons to be read in different groups of three. If a deletion is combined with an addition to produce a double-mutant gene, the frameshift occurs only in the region between the two mutants. If this region is small enough or does not contain coding for vital amino acids, the function of the gene may be restored. Two deletions or two insertions combined will not restore the reading frame. However, Crick and his colleagues found that the combination of three additions or three deletions did restore gene function. This finding led to the conclusion that the genetic code was triplet, because a triplet code would be put back into the reading frame by three additions or three deletions (fig. 11.29).

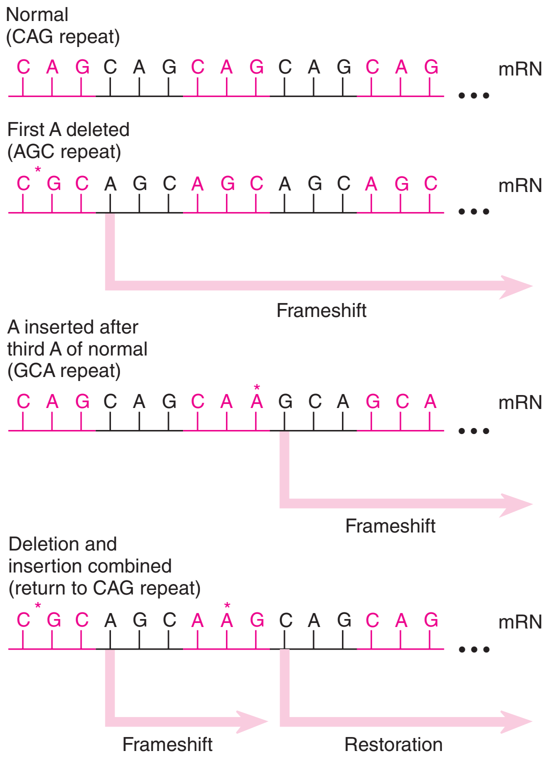

Normal
(CAG repeat)
C A G C A G C A G C A G C A G
mRNA
First A deleted
(AGC repeat)
C * G C A G C A G C A G C A G C
mRNA
Frameshift
A inserted after
third A of normal
(GCA repeat)
C A G C A G C A * A G C A G C A
mRNA
Frameshift
Deletion and
insertion combined
(return to CAG repeat)
C * G C A G C A * A G C A G C A G
mRNA
Frameshift
Restoration

**Figure 11.28** Frameshift mutations in a gene result from the addition or deletion of one or several nucleotides (any number other than a multiple of three) in the DNA. The messenger RNA shown here normally has a CAG repeat. A single-base deletion shifts the three-base reading frame to a series of AGC repeats. A later insertion restores the reading frame. *Asterisks* (*) indicate points of deletion or insertion.

### Overlap and Punctuation in the Code

Questions still remained: was the code overlapping? Did it have punctuation? Several logical arguments favored a no-punctuation, nonoverlapping model (fig. 11.30). An overlapping code would be subject to two restrictions. First, a change in one base (a mutation) could affect more than one codon and thus affect more than one amino acid. But studies of amino acid sequences almost always showed that only one amino acid was changed, which argued against codon overlap. Second, certain restrictions affected which amino acids occurred next to each other in proteins. For example, the amino acid UUU coded could never be adjacent to the amino acid coded by AAA because one or both (depending on the number of bases overlapped) of the overlap codons UUA and UAA would always insert other amino acids between them. Overlap, then, seemed to be ruled out since every amino acid appears next to every other amino acid in one protein or another.

Punctuation between codons was also tentatively ruled out. The messenger RNA in the tobacco necrosis satellite virus has just about enough codons to specify its coat protein with no room left for a punctuating base or bases between each codon.

## Breaking the Code

Once geneticists had figured out that the genetic code is in nonoverlapping triplets, they turned their attention to the sixty-four codons. They wondered which amino acid, for example, does ACU specify? The work was done in two stages. In the first stage, M. W. Nirenberg, S. Ochoa, and their colleagues made long artificial messenger RNAs and determined which amino acids these messenger RNAs incorporated into protein. In the second stage, specific triplet RNA sequences were synthesized. The researchers then determined the amino acid-transfer RNA complex that was bound by each sequence.

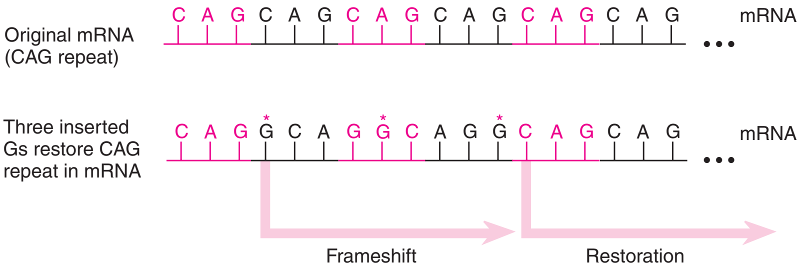
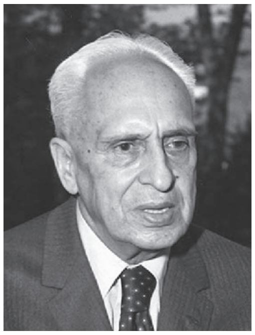
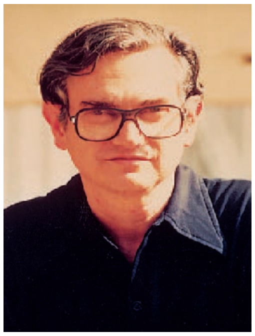

C A G C A G C A G C A G C A G C A G
mRNA
Original mRNA
(CAG repeat)
Three inserted
Gs restore CAG
repeat in mRNA
C A G G* C A G G* C A G G* C A G C A G
mRNA
Frameshift
Restoration

**Figure 11.29** The coding frame of CAG repeats is first shifted and then restored by three additions (insertions). *Asterisks* (*) indicate insertions.



Severo Ochoa (1905–1993). (Courtesy of Dr. Severo Ochoa.)



Marshall W. Nirenberg (1934– ). (Courtesy of Dr. Marshall W. Nirenberg.)

### *Synthetic Messenger RNAs*

The ability to synthesize long-chain messenger RNAs resulted from the 1955 discovery of M. Grunberg-Manago and Ochoa of the enzyme **polynucleotide phosphorylase,** which joins diphosphate nucleotides into long-chain, single-stranded polynucleotides. Unlike a polymerase, polynucleotide phosphorylase does not need a primer on which to act. This enzyme is found in all bacteria. (Its main function in the cell is probably the reverse of its use here. It most likely serves as an exonuclease, degrading messenger RNA.) In 1961, Nirenberg and J. H. Matthei added artificially formed RNA polynucleotides of known composition to an *E. coli* ribosomal system and looked for the incorporation of amino acids into proteins.

The system just described is called a **cell-free system,** a mixture primarily of the cytoplasmic components of cells, such as *E. coli,* but missing nucleic acids and membrane components. These systems are relatively easy to create by disrupting and then fractionating whole cells. The systems hold the advantage of containing virtually all the components needed for protein synthesis except the messenger RNAs. Their disadvantages are that they are relatively short-lived (several hours) and are relatively inefficient in translation. However, an added benefit to the *E. coli* cell-free system is that it will translate, albeit inefficiently, RNAs that normally are not translated in vivo because they lack translation initiation signals. This feature allowed these scientists to use artificial messenger RNAs that contained no Shine-Dalgarno sequence for ribosomal binding.

Nirenberg and Matthei found that when the enzyme polynucleotide phosphorylase in the *E. coli* cell-free system made uridine diphosphates into a poly-U messenger RNA, phenylalanine residues were incorporated into a polypeptide. Thus, the first code word established was UUU for phenylalanine. Nirenberg and Ochoa and their associates continued the work. They found that AAA was the code word for lysine, CCC was the code word for proline, and GGG was the code word for glycine.

They then made synthetic messenger RNAs by using mixtures of the various diphosphate nucleotides in known proportions. Table 11.3 gives an example. From their experiments, it was possible to determine the bases used in many of the code words, but not their specific order. For example, cysteine, leucine, and valine are all coded by two Us and a G, but the experiment could not sort out the order of these bases (5′-UUG-3′, 5′-UGU-3′, or 5′-GUU-3′) for any one of them. Determining the order required an extra step in sophistication—that is, being able to synthesize known trinucleotides.

### *Synthetic Codons*

Once trinucleotides of known composition could be manufactured, Nirenberg and P. Leder in 1964 developed a "binding assay." They found that isolated *E. coli* ribosomes, in the presence of high-molarity magnesium chlo-

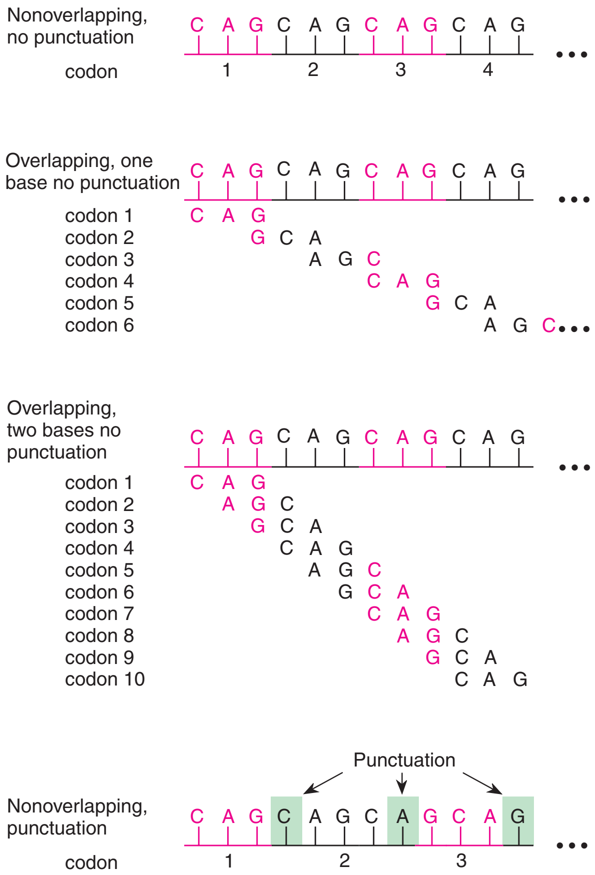

Nonoverlapping, no punctuation
C A G C A G C A G C A G
codon
1
2
3
4
Overlapping, one base no punctuation
C A G C A G C A G C A G
codon 1
C A G
codon 2
G C A
codon 3
A G C
codon 4
C A G
codon 5
G C A
codon 6
A G C
Overlapping, two bases no punctuation
C A G C A G C A G C A G
codon 1
C A G
codon 2
A G C
codon 3
G C A
codon 4
C A G
codon 5
A G C
codon 6
G C A
codon 7
C A G
codon 8
A G C
codon 9
G C A
codon 10
C A G
Punctuation
Nonoverlapping, punctuation
C A G C A G C A G C A G
codon
1
2
3

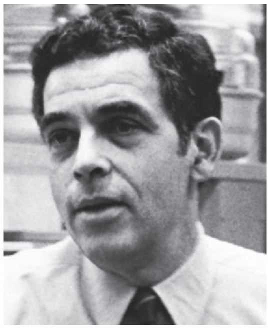
**Figure 11.30** The genetic code is read as a nonoverlapping code with no punctuation (*top*). Before that was proven, it was suggested that the code could overlap by one or two bases (*middle*) or have noncoded bases (punctuation) between code words (*bottom*).

**Table 11.3** **Structure of Artificial mRNA Made by Randomly Assembling Uracil- and Guanine-Containing Ribose Diphosphate Nucleotides with a Ratio of 5U:lG**

| Codon | Frequency of Occurrence |
|---|---|
| UUU | $(5/6)^3 = 0.58$ |
| UUG | $(5/6)^2(1/6) = 0.12$ |
| UGU | $(5/6)^2(1/6) = 0.12$ |
| GUU | $(5/6)^2(1/6) = 0.12$ |
| UGG | $(5/6)(1/6)^2 = 0.02$ |
| GUG | $(5/6)(1/6)^2 = 0.02$ |
| GGU | $(5/6)(1/6)^2 = 0.02$ |
| GGG | $(1/6)^3 = 0.005$ |

ride, could bind trinucleotides as if they were messenger RNAs. Also bound was the transfer RNA that carried the anticodon complementary to the trinucleotide. It was thus possible, using radioactive amino acids, to determine which messenger RNA trinucleotide coded for a particular amino acid. A given synthetic trinucleotide was mixed with ribosomes and aminoacyl-tRNAs, including one radioactively labeled amino acid. The reaction mixture was passed over a filter that would allow everything except the large trinucleotide + ribosome + aminoacyl-tRNA complex to pass through. If the radioactivity passed through the filter, it meant that the radioactive amino acid was not associated with the ribosome. The experiment was then repeated with another labeled amino acid. When the radioactivity appeared on the filter, the investigators knew that the amino acid was affiliated with the ribosome. Thus, that amino acid was coded by the selected trinucleotide codon. In other words, the radioactive amino acid was attached to a transfer RNA whose anticodon was complementary to the trinucleotide codon and thus bound at the ribosome.

Figure 11.31 shows an example. In the figure, the trinucleotide is 5′-CUG-3′. The transfer RNA with the anticodon 3′-GAC-5′ is charged with leucine. The mixture is passed through a filter. If threonine, or any other amino acid except leucine, is radioactive, the radioactivity passes through the filter. When the experiment is repeated with radioactive leucine, the leucine, and hence the radioactivity, is trapped by the filter. In a short period of time, all of the codons were deciphered (table 11.4).


Phillip Leder (1934– ).
(Courtesy of Dr. Phillip Leder.)

## *Wobble Hypothesis*

The genetic code is a **degenerate code,** meaning that a given amino acid may have more than one codon. As you can see from table 11.4, eight of the sixteen boxes contain just one amino acid per box. (A box is determined by the first and second positions; e.g., the UUX box, in which X is any of the four bases.) Therefore, for these eight amino acids, the codon need only be read in the first two positions because the same amino acid will be represented regardless of the third base of the codon. These eight groups of codons are termed **unmixed families** of codons. An unmixed family is the four codons beginning with the same two bases that specify a single amino acid. For example, the codon family GUX codes for valine. **Mixed families** code for two amino acids or for stop signals and one or two amino acids.

Six of the mixed-family boxes are split in half so that the codons are differentiated by the presence of a purine or a pyrimidine in the third base. For example, CAU and

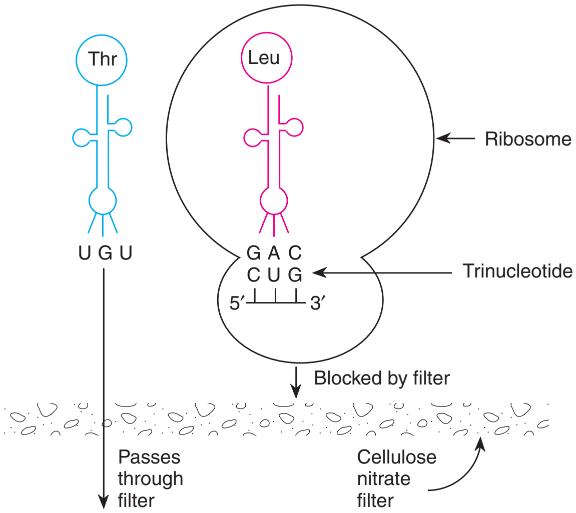


**Figure 11.31** The binding assay determines the amino acid associated with a given trinucleotide codon. Transfer RNAs with noncomplementary codons pass through the membrane. Transfer RNAs with anticodons complementary to the trinucleotide bind to the ribosome and do not pass through the filter. When the transfer RNA is charged with a radioactive amino acid, the radioactivity is trapped on the filter.

**Table 11.4 The Genetic Code**

| | Second Position | | | | |
|---|---|---|---|---|---|
| **First Position (5′ End)** | **U** | **C** | **A** | **G** | **Third Position (3′ End)** |
| **U** | Phe | Ser | Tyr | Cys | **U** |
| | Phe | Ser | Tyr | Cys | **C** |
| | Leu | Ser | *stop* | *stop* | **A** |
| | Leu | Ser | *stop* | Trp | **G** |
| **C** | Leu | Pro | His | Arg | **U** |
| | Leu | Pro | His | Arg | **C** |
| | Leu | Pro | Gln | Arg | **A** |
| | Leu | Pro | Gln | Arg | **G** |
| **A** | Ile | Thr | Asn | Ser | **U** |
| | Ile | Thr | Asn | Ser | **C** |
| | Ile | Thr | Lys | Arg | **A** |
| | Met (*start*) | Thr | Lys | Arg | **G** |
| **G** | Val | Ala | Asp | Gly | **U** |
| | Val | Ala | Asp | Gly | **C** |
| | Val | Ala | Glu | Gly | **A** |
| | Val | Ala | Glu | Gly | **G** |

CAC both code for histidine; in both, the third base, U (uracil) or C (cytosine), is a pyrimidine. Only two of the families of codons are split differently.

The lesser importance of the third position in the genetic code ties in with two facts about transfer RNAs. First, although there would seem to be a need for sixty-two transfer RNAs—since there are sixty-one codons specifying amino acids and an additional codon for initiation—there are actually only about fifty different transfer RNAs in an *E. coli* cell. Second, a rare base such as inosine can appear in the anticodon, usually in the position that is complementary to the third position of the codon. These two facts lead researchers to believe that some kind of conservation of transfer RNAs is occurring and that rare bases may be involved.

We should mention, to avoid confusion, that both messenger RNA and transfer RNA bases are usually numbered from the 5′ side. Thus, the number-one base of the codon is complementary to the number-three base of the anticodon (fig. 11.32). Thus, the codon base of lesser importance is the number-three base, whereas its complement in the anticodon is the number-one base.

Since the first position of the anticodon (5′) is not as constrained as the other two positions, a given base at that position may be able to pair with any of several bases in the third position of the codon. Crick characterized this ability as **wobble** (fig. 11.33). Table 11.5 shows the possible pairings that would produce a transfer RNA system compatible with the known code. For example, if an isoleucine transfer RNA has the anticodon 3′-UAI-5′, it is compatible with the three codons for that amino acid (see table 11.4): 5′-AUU-3′, 5′-AUC-3′, and 5′-AUA-3′. That is, inosine in the first (5′) position of the anticodon can recognize U, C, or A in the third (3′) position of the codon, and thus one transfer RNA complements all three codons for isoleucine.

## *Universality of the Genetic Code*

Until 1979, scientists concluded that the genetic code was universal. That is, the codon dictionary (see table 11.4) was the same for *E. coli,* human beings, and oak trees, as well as all other species studied up to that time. The universality of the code was demonstrated, for example, by taking the ribosomes and messenger RNA from rabbit reticulocytes and mixing them with the aminoacyl-tRNAs and other translational components of *E. coli.* Rabbit hemoglobin was synthesized.

In 1979 and 1980, however, researchers noted discrepancies when sequencing mitochondrial genes for struc-

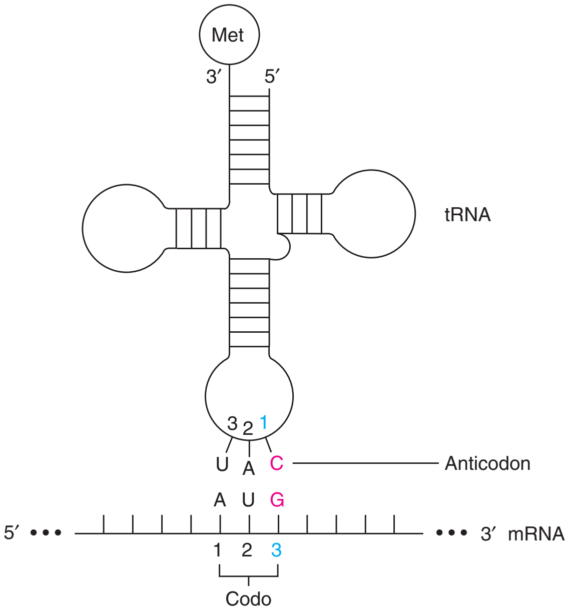


**Figure 11.32** Codon and anticodon base positions are numbered from the 5′ end. The 3′ position in the codon (5′ in the anticodon) is the wobble base.

tural proteins (see chapters 13 and 17). It was discovered that there were two kinds of deviations from universality in the way mitochondrial transfer RNAs read the code. First, fewer transfer RNAs were needed to read the code. Second, there were several instances in which the mitochondrial and cellular systems interpreted a codon differently.

According to Crick's wobble rules (see table 11.5), thirty-two transfer RNAs (including one for initiation) can complement all sixty-one nonterminating codons. Unmixed families require two transfer RNAs, and mixed families require one, two, or three transfer RNAs, depending on the family. The yeast mitochondrial coding system apparently needs only twenty-four transfer RNAs. The reduction in numbers is accomplished primarily by having only one transfer RNA recognize each unmixed family (table 11.6; cf. table 11.4). Because mitochondrial transfer RNAs for unmixed families of codons have a U in the first (wobble) position of the anticodon, apparently, given the structure of the mitochondrial transfer RNAs, the U can pair with U, C, A, or G. Presumably, evolutionary pressure has minimized the number of transfer RNA genes in the DNA of the mitochondrion, in keeping with its small size. Reduction from thirty-two to twenty-four is a 25% savings. (Recent evidence suggests that mammalian mitochondria may need only twenty-two transfer RNAs.)

**Table 11.5** **Pairing Combinations at the Third Codon Position**

| Number-one Base in tRNA (5′ End) | Number-three Base in mRNA (3′ End) |
|---|---|
| G | U *or* C |
| C | G |
| A | U |
| U | A *or* G |
| I | A, U, *or* C |

It has also been found that yeast mitochondria read the CUX family as threonine rather than as leucine (tables 11.4 and 11.6) and the terminator UGA (opal) as tryptophan rather than as termination. However, there appear to be differences among different groups of organisms reading the CUX family. Human and *Neurospora* mitochondria appear to read the CUX codons as leucine, just as cellular systems do. Of the groups so far analyzed, only yeast reads the CUX family as threonine. Similarly, human and *Drosophila* mitochondria read AGA and AGG as stop signals rather than as arginine (table 11.7).

In 1985, it was discovered that *Paramecium* species read the UAA and UAG stop codons as glutamine within the cell. In addition, a prokaryote (*Mycoplasma capricolum*) reads UGA as tryptophan. We do not yet know how general this finding is: scientists have scrutinized the genetic code of very few species. We can thus conclude that the genetic code seems to have universal tendencies among prokaryotes, eukaryotes, and viruses. Mitochondria, however, read the code slightly differently: different wobble rules apply, and mitochondria and cells read at least one terminator and one unmixed family of codons differently. Also, the mitochondrial discrepancies are not universal among all types of mitochondria. Further work, involving the sequencing of more mitochondrial DNAs, should elucidate the pattern of discrepancies among the mitochondria of diverse species. We also now know that not every organism reads all codons in the same way. Ciliated protozoa and a mycoplasma read some stop signals as coding for amino acids. Nuclear variants are known in the following codons: CUG, AUA, UAA, UAG, UGA, CGG, and AGA. Mitochondrial variants are known in CUX, AUA, UAA, UAG, AAA, UGA, CGX, AGA, and AGG.

One other type of variation of codon reading occurs: *site-specific variation,* in which the interpretation of a codon depends on its specific location. We are already familiar with the fact that GUG and, rarely, UUG can serve as prokaryotic initiation codons. This means that they are recognized by $\text{tRNA}_\text{f}^\text{Met}$. However, they are not recognized by $\text{tRNA}_\text{m}^\text{Met}$ (i.e., GUG and UUG are not misread internally in messenger RNAs). In some cases, two of the termination codons (UGA and UAG, but not UAA) are misinterpreted as codons for amino acids. That is, termination will not occur at the normal place, resulting in a longer-than-usual protein. In some cases, these "read-through" proteins are vital—the

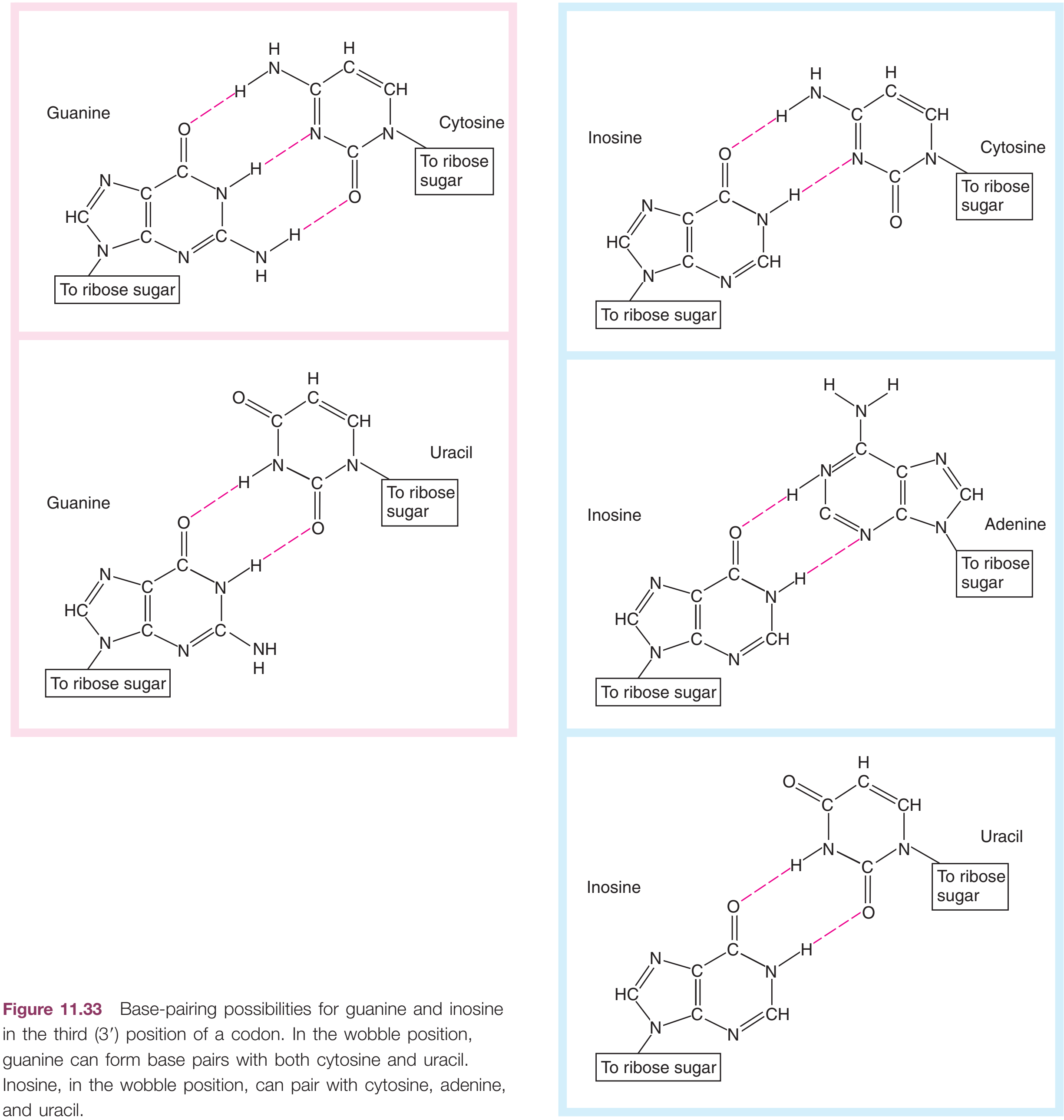

Guanine
Cytosine
To ribose sugar
Guanine
Uracil
To ribose sugar
Inosine
Cytosine
To ribose sugar
Inosine
Adenine
To ribose sugar
Inosine
Uracil
To ribose sugar

**Figure 11.33** Base-pairing possibilities for guanine and inosine in the third (3′) position of a codon. In the wobble position, guanine can form base pairs with both cytosine and uracil. Inosine, in the wobble position, can pair with cytosine, adenine, and uracil.

organism depends on their existence. For example, in the phage Qβ, the coat-protein gene is read through about 2% of the time. Without this small number of read-through proteins, the phage coat cannot be constructed properly.

One last example of site-specific variation involves the amino acid selenocysteine (cysteine with a selenium atom replacing the sulfur; see fig. 11.1). Although many proteins have unusual amino acids, almost all are due to posttranslational modifications of normal amino acids. However, the amino acid selenocysteine is inserted directly into some proteins, such as formate dehydrogenase in *E. coli,* which has selenium in its active site. Selenocysteine is inserted into the protein by a novel transfer RNA that recognizes the termination codon, UGA, if that codon is involved in a particular stem-loop secondary structure in the messenger RNA. The selenocysteine transfer RNA is originally charged with a serine that is then modified to a selenocysteine. In addition to the

**Table 11.6** **The Genetic Code Dictionary of Yeast Mitochondria***

| | Second Position | | | | |
|---|---|---|---|---|---|
| **First Position (5′ End)** | **U** | **C** | **A** | **G** | **Third Position (3′ End)** |
| **U** | Phe AAG | Ser AGU | Tyr AUG | Cys ACG | **U** |
| | | | | | **C** |
| | Leu AAU | | *stop* | Trp ACU | **A** |
| | | | | | **G** |
| **C** | Thr GAU | Pro GGU | His GUG | Arg GCA | **U** |
| | | | | | **C** |
| | | | Gln GUU | | **A** |
| | | | | | **G** |
| **A** | Ile UAG | Thr UGU | Asn UUG | Ser UCG | **U** |
| | | | | | **C** |
| | Met UAC | | Lys UUU | Arg UCU | **A** |
| | | | | | **G** |
| **G** | Val CAU | Ala CGU | Asp CUG | Gly CCU | **U** |
| | | | | | **C** |
| | | | Glu CUU | | **A** |
| | | | | | **G** |

*Source:* Data from S. Bonitz, et al., "Codon recognition rules in yeast mitochondria," *Proceedings of the National Academy of Sciences* 77:3167–70, 1980.

* Anticodons (3′ → 5′) are given within boxes. (The ACU Trp anticodon is predicted.)

**Table 11.7** **Common and Alternative Meanings of Codons**

| Codon | General Meaning | Alternative Meaning |
|---|---|---|
| CUX | Leu | Thr in yeast mitochondria |
| AUA | Ile | Met in mitochondria of yeast, *Drosophila,* and vertebrates |
| UGA | *Stop* | Trp in mycoplasmas and mitochondria other than higher plants |
| AGA/AGG | Arg | *Stop* in mitochondria of yeast and vertebrates |
| | | Ser in mitochondria of *Drosophila* |
| CGG | Arg | Trp in mitochondria of higher plants |
| UAA/UAG | *Stop* | Gln in ciliated protozoa |
| UAG | *Stop* | Ala or Leu in mitochondria of some higher plants |

stem-loop structure 3′ (downstream) from the amber codon (UAG), a selenocysteine elongation factor (SELB) is also needed at the ribosome. This same mechanism may occur in eukaryotes, but not all of the components have yet been identified.

## *Evolution of the Genetic Code*

It has been theorized that the genetic code has wobble in it because it originally arose from a code in which only the first two bases were needed for the small number of amino acids in use several billion years ago. As new amino acids with useful properties became available, they were incorporated into proteins by a code modified by the third base, albeit with less specificity. This view has support from the fact that codons starting with the same nucleotide come from the same biosynthetic pathway. This indicates that in early evolution, as biosynthetic pathways were extended to new amino acids, the newcomers were incorporated by use of the second and third bases of the code.

However, the question remains as to whether the genetic code is highly evolved or just a "frozen accident." In other words, is there a relationship between the codons and the amino acids they code for, or is the code just one of many random possibilities? Recent computer simulations of random codes indicate that the current genetic code is far outside the range of random in its ability to protect the organism from mutation. This suggests that the genetic code is not a frozen accident, but rather is highly evolved. Numerous examples in the current code support this view.

For example, in the unmixed codon family 5′-CUX-3′, any mutation in the third position produces another codon for the same amino acid. Wobble in the third position and codon arrangement ensures that less than half of the mutations in the third codon position result in the specification of a different amino acid.

There are also patterns in the genetic code in which the mutation of one codon to another results in an amino acid of similar properties. A high probability exists that such a mutation will produce a functional protein. All the codons with U as the middle base, for example, are for amino acids that are hydrophobic (phenylalanine, leucine, isoleucine, methionine, and valine). Mutation in the first or third positions for any of these codons still codes a hydrophobic amino acid. Both of the two negatively charged amino acids, aspartic acid and glutamic acid, have codons that start with GA. All of the aromatic amino acids—phenylalanine, tyrosine, and tryptophan (see fig. 11.1)—have codons that begin with uracil. Such patterns minimize the negative effects of mutation.

This chapter completes the discussion of the mechanics of gene expression. The next chapter deals with recombinant DNA technology, followed by several chapters concerned with the control of gene expression in both prokaryotes and eukaryotes.

## SUMMARY

**STUDY OBJECTIVE 1:** To study the mechanism of protein biosynthesis, in which organisms, using the information in DNA, string together amino acids to form proteins 281–303

A charged transfer RNA has an anticodon at one end and a specific amino acid at the other end. The transfer RNAs are charged with the proper amino acid by aminoacyl-tRNA synthetase enzymes that incorporate the energy of ATP into amino acid-tRNA bonds. Hence, no additional source of energy is needed during peptide bond formation. During protein synthesis, the translation apparatus at the ribosome recognizes the transfer RNA. Through complementarity, the anticodon pairs with a messenger RNA codon.

An initiation complex forms at the start of translation. In prokaryotes, this complex consists of the messenger RNA, the 30S subunit of the ribosome, the initiator transfer RNA with N-formyl methionine (fMet-$\text{tRNA}_\text{f}^\text{Met}$), and the initiation factors IF1, IF2, and IF3. The 50S ribosomal subunit is then added and A and P sites form in the resulting 70S ribosome. The charged N-formyl methionine transfer RNA is in the P site. A GTP is hydrolyzed, and the initiation factors are released.

A transfer RNA enters the A site, which requires the involvement of elongation factors EF-Ts and EF-Tu (in *E. coli*). At least one GTP hydrolysis releases the elongation factor, EF-Tu, which had originally brought the charged transfer RNA to the ribosome. Peptidyl transferase, which appears to be a ribozymic component of the 50S ribosomal subunit, transfers the amino acid from the transfer RNA in the P site to the amino end of the amino acid on the transfer RNA in the A site.

With the help of elongation factor G (EF-G), the ribosome translocates in relation to the messenger RNA. The depleted transfer RNA is moved from the P site to the E site, where it is released; the transfer RNA with the growing peptide is moved into the P site. EF-G is then released. Elongation and translocation continue until a nonsense codon enters the A site. With the aid of the release factors RF1 and RF2, the protein is released, and the messenger RNA-ribosome complex dissociates. Eukaryotes have slightly more complex processes involving several more proteins.

Proteins pass through membranes with the help of a signal peptide synthesized at their N-terminal ends. Proteins fold into their final, functional configurations with the help of molecular chaperones, proteins that aid the folding process.

Many antibiotics interfere with translation in prokaryotes. Puromycin, streptomycin, tetracycline, and chloramphenicol all act at the ribosome. Studying the mode of action of these antibiotics has provided insights into the mechanism of the translation process.

**STUDY OBJECTIVE 2:** To examine the genetic code 304–312

The genetic code was first assumed to be triplet because of logical arguments regarding the minimum size of codons. With his work on deletion and insertion mutants, Crick provided evidence that the code was triplet. Part of the code was worked out initially with the synthesis of long, artificial messenger RNAs and then the synthesis of specific trinucleotide codons. Crick's wobble hypothesis accounts for the fact that fewer than sixty-one transfer RNAs can read the entire genetic code. Fewer transfer RNAs are needed because additional complementary base pairings occur in the third position (3′) of the codon.

The rule of universality of the genetic code has to be modified in light of findings regarding mitochondrial transfer RNAs; only twenty-four are needed to read the code. In addition, some sense codons are interpreted differently in mitochondrial systems; some nonmitochondrial systems read stop codons differently (a mycoplasma and ciliated protozoan); and some site-specific variation in codon reading also occurs. The structure of the code in both cells and mitochondria seems to protect the cell against a good deal of potential mutation.

## SOLVED PROBLEMS

**PROBLEM 1:** What is the energy requirement of protein biosynthesis?

***Answer:*** The cost of adding one amino acid to a growing polypeptide is four or five high-energy bonds: two from an ATP during the charging of the transfer RNA, and two or three from the hydrolysis of GTPs during transfer RNA binding to the A site of the ribosome and during translocation. Thus, for an average protein of three hundred amino acids, there is a cost of 1,200 to 1,500 high-energy bonds.

**PROBLEM 2:** What are the start and stop signals of translation?

***Answer:*** Once a messenger RNA is attached at the ribosome, the start signal is the methionine initiation codon (usually AUG), whereas the stop signal is one of the three nonsense codons (UAA, UAG, and UGA). Binding to the ribosome in order to position the messenger RNA in relation to the A and P sites differs in prokaryotes and eukaryotes. In prokaryotes, the Shine-Dalgarno sequence allows the 16S ribosomal RNA and the messenger RNA to form hydrogen bonds, locating the beginning of the messenger RNA at the ribosome. In eukaryotes, the 5′ cap is usually recognized by the ribosome, and the ribosome then proceeds to scan the messenger RNA for the initiation codon.

**PROBLEM 3:** What amino acids could replace methionine if a one-base mutation occurred?

***Answer:*** The codon for methionine (internal as well as initiation) is AUG. If the A is replaced, we would get UUG (Leu), CUG (Leu), or GUG (Val); if the U is replaced, we would get AAG (Lys), ACG (Thr), or AGG (Arg); and if the G is replaced, we would get AUA (Ile), AUU (Ile), or AUC (Ile). Hence, a one-base change in the codon for methionine could result in any of six different amino acids.

## EXERCISES AND PROBLEMS*

**INFORMATION TRANSFER**

1. Given the following end part of a gene, which will be transcribed and then translated into a pentapeptide, provide the base sequence for its messenger RNA. Give the anticodons on the transfer RNAs by making use of wobble rules. What amino acids are incorporated? Draw the actual structure of the pentapeptide.

   3′-TACAATGGCCCTTTTATC-5′
   5′-ATGTTACCGGGAAAATAG-3′

2. Give an alternative translation mechanism that would require only one transfer RNA site on the ribosome.
3. Draw the details of a moment in time at the ribosome during the translation of the messenger RNA produced in problem 1. Include in the diagram the ribosomal sites, the transfer RNAs, and the various nonribosomal proteins involved.
4. How do prokaryotic and eukaryotic ribosomes recognize the 5′ end of messenger RNAs? Could eukaryotic messenger RNAs be polycistronic?
5. How many aminoacyl-tRNA synthetases are there? What do they use for recognition signals?
6. What are the similarities and differences among the three nonsense codons? Using the wobble rules, what are their theoretical anticodons?
7. Describe an experiment that demonstrates that the transfer RNA, and not its amino acid, is recognized at the ribosome during translation.
8. Other than the antibiotics named in the chapter, suggest five "theoretical" antibiotics that could interfere with the prokaryotic translation process.
9. How many single-base deletions are required to restore the reading frame of a messenger RNA? Give an example.
10. A "nonsense mutation" is one in which a codon for an amino acid changes to one for chain termination. Give an example. What are its consequences?
11. The reverse situation to problem 10 is a mutation from a nonsense codon to a codon for an amino acid. Give an example. What are its consequences?
12. What are the consequences when an internal methionine codon recognizes a prokaryotic initiation transfer RNA?
13. What role does EF-Ts play in elongation? EF-Tu? What are their eukaryotic equivalents?
14. What roles do RF1 and RF2 play in chain termination? What are their eukaryotic equivalents?
15. What is a signal peptide? What role does it play in eukaryotes? What is its fate?
16. Why doesn't puromycin disrupt eukaryotic translation?

---

*Answers to selected exercises and problems are on page A-12.

**17.** A peptide, fifteen amino acids long, is digested by two methods, and each segment is sequenced according to the Edman degradation technique (see box 11.1). The fifteen amino acids are denoted by the letters A through O, with F as the N-terminal amino acid. If the segments are as follows, what is the sequence of the original peptide?

Method 1: CABHLN; FGKI; OEDJM

Method 2: KICAB; JM; FG; HLNOED

**18.** In human hemoglobin, the β chain is 146 amino acids long. What is the minimum length of RNA needed to make this protein?

**19.** Part of a DNA strand to be transcribed has the following sequence:

3′-TACTAACTTACGCTCGCCTCA-5′

**a.** What is the sequence of RNA transcribed from this part of the strand?

**b.** What sequence of amino acids does the RNA produce?

**THE GENETIC CODE**

**20.** If DNA contained only the bases cytosine and guanine, how long would a code word have to be? How could we tell if this DNA were double-stranded?

**21.** If an artificial messenger RNA contains two parts uracil to one of cytosine, name the amino acids and the proportions in which they should be incorporated into protein.

**22.** What would be proved or disproved if an organism were discovered that did not follow any of the rules of the codon dictionary? Would we expect organisms from another galaxy (if they exist) to use our codon dictionary?

**23.** What would the genetic code dictionary (see table 11.4) look like if wobble occurred in the second position rather than the third (i.e., if an unmixed family of codons were of the form GXU)?

**24.** In experiments using repeating polymers, $(GCGC)_n$ incorporates alanine and arginine into polypeptides, and $(CGGCGG)_n$ incorporates arginine, glycine, and alanine. What codon can probably be assigned to glycine?

**25.** If poly-G is used as a messenger RNA in an incorporation experiment, glycine is incorporated into a polypeptide. If poly-C is used, proline is incorporated. If both poly-G and poly-C are used, no amino acids are incorporated into protein. Why?

**26.** A protein has leucine at a particular position. If the codon for leucine is CUC, how many different amino acids might appear as the result of a single-base substitution?

**27.** Polymers of $(GUA)_n$ result in the incorporation of only two different amino acids rather three, as for most other three-base polymers. Why?

**28.** The sixth amino acid in the β chain of normal human hemoglobin is glutamate. Two different mutations of this codon substitute valine and lysine. What is the likely codon for glutamate?

**29.** A normal protein has the following C-terminal amino acid sequence: *ser-thr-lys-leu*-COOH. A mutant is isolated with the following sequence: *ser-thr-lys-leu-leu-phe-arg*-COOH. What has probably happened to produce the mutant protein?

**30.** A segment of a normal protein and three different mutants appears as follows:

| | |
|---|---|
| normal | _____gly-ala-ser-his-cys-leu-phe_____ |
| mutant 1 | _____gly-ala-ser-his |
| mutant 2 | _____gly-ala-ser-leu-cys-leu-phe_____ |
| mutant 3 | _____gly-val-ala-ile-ala-ser_____ |

What is the probable sequence of bases in the normal RNA?

**31.** A normal protein has histidine in a given position. Four mutants are isolated and determined to have either tyrosine, glutamine, proline, or leucine in place of histidine. What are the possible codon assignments, and what codon is probably used for histidine?

## CRITICAL THINKING QUESTIONS

**1.** How do transcriptional and translational signals interact?

**2.** Why is the E site necessary in translation?

*Suggested Readings for chapter 11 are on page B-8.*

# 12

# DNA

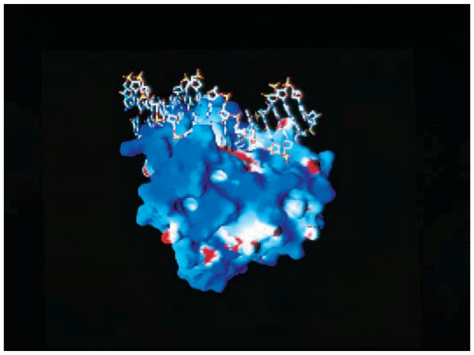
## *Its Mutation, Repair, and Recombination*

## STUDY OBJECTIVES

1. To look at the nature of mutation in prokaryotes 316
2. To analyze functional and structural allelism and examine the mapping of mutant sites within a gene 317
3. To verify the colinearity of gene and protein 324
4. To study mutagenesis 325
5. To investigate the processes of DNA repair and recombination 339

## STUDY OUTLINE

**Mutation** 316
  Fluctuation Test 316
  Genetic Fine Structure 317
  Colinearity 324
  Spontaneous Versus Induced Mutation 325
  Mutation Rates 326
  Point Mutations 326
  Spontaneous Mutagenesis 328
  Chemical Mutagenesis 330
  Misalignment Mutagenesis 337
  Intergenic Suppression 337
  Mutator and Antimutator Mutations 338
**DNA Repair** 339
  Damage Reversal 340
  Excision Repair 340
  Double-Strand Break Repair 344
  Postreplicative Repair 344
**Recombination** 347
  Double-Strand Break Model of Recombination 347
  Bacterial Recombination 349
  Hybrid DNA 351
*Summary* 352
*Solved Problems* 354
*Exercises and Problems* 354
*Critical Thinking Questions* 356
**Box 12.1 The Ames Test for Carcinogens** 332
**Box 12.2 In Vitro Site-Directed Mutagenesis** 333
**Box 12.3 Adaptive Mutation** 339



Computer-generated space-filling model of a DNA enzyme repairing damaged DNA.

(© James King-Holmes/SPL/Photo Researchers, Inc.)

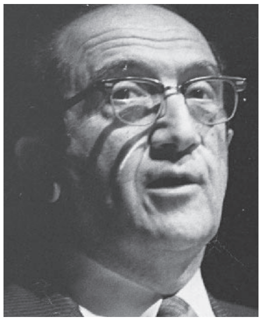
The mutation, repair, and recombination of DNA are treated together in this chapter because the three processes have much in common. The physical alteration of DNA is involved in each; repair and recombination share some of the same enzymes. We progress from mutation—the change in DNA—to repair of damaged DNA, and, finally, to recombination, the new arrangement of pieces of DNA.

## MUTATION

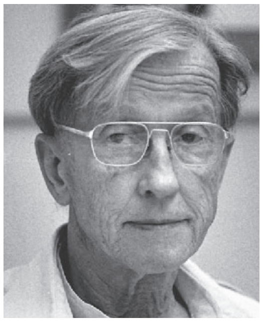
The concept of mutation (a term coined by de Vries, a rediscoverer of Mendel) is pervasive in genetics. **Mutation** is both the process by which a gene (or chromosome) changes structurally and the end result of that process. Without alternative forms of genes, the biological diversity that exists today could not have evolved. Without alternative forms of genes, it would have been virtually impossible for geneticists to determine which of an organism's characteristics are genetically controlled. Studies of mutation provided the background for our current knowledge in genetics.

### Fluctuation Test

In 1943, Salvador Luria and Max Delbrück published a paper entitled "Mutations of Bacteria from Virus Sensitivity to Virus Resistance." This paper ushered in the era of bacterial genetics by demonstrating that the phenotypic variants found in bacteria are actually attributable to mutations rather than to induced physiological changes. Very little work had previously been done in bacterial genetics because of the feeling that bacteria did not have "normal" genetic systems like the systems of fruit flies and corn. Rather, bacteria were believed to respond to environmental change by physiological adaptation, a non-Darwinian view. As Luria said, bacteriology remained "the last stronghold of Lamarckism" (the belief that acquired characteristics are inherited).



Salvador E. Luria (1912–1991). (Courtesy of Dr. S. E. Luria.)



Max Delbrück (1906–1981). (Courtesy of Dr. Max Delbrück.)

#### What Causes Genetic Variation?

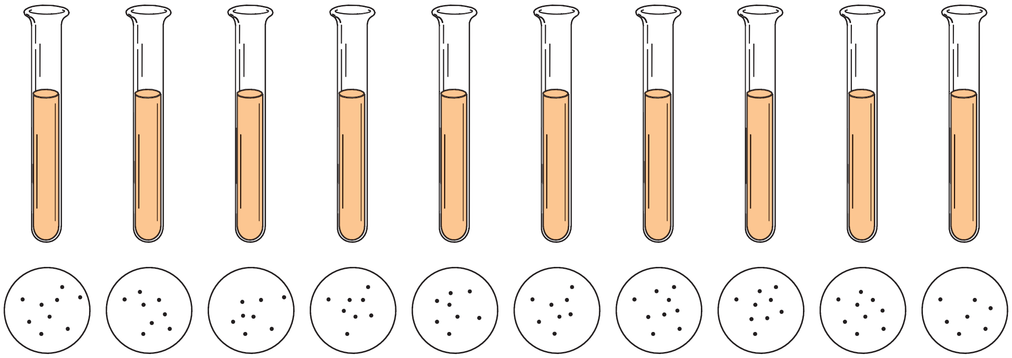
Luria and Delbrück studied the Ton$^r$ (phage T1-resistant) mutants of a normal Ton$^s$ (phage T1-sensitive) *Escherichia coli* strain. They used an enrichment experiment, as described in chapter 7, wherein a petri plate is spread with *E. coli* bacteria and T1 phages. Normally, no bacterial colonies grow on the plate: all the bacteria are lysed. However, if one of the bacterial cells is resistant to T1 phages, it produces a bacterial colony, and all descendants of this colony are T1 resistant. There are two possible explanations for the appearance of T1-resistant colonies:

1. Any *E. coli* cell may be induced to be resistant to phage T1, but only a very small number actually are. That is, all cells are genetically identical, each with a very low probability of exhibiting resistance in the presence of T1 phages. When resistance is induced, the cell and its progeny remain resistant.
2. In the culture, a small number of *E. coli* cells exist that are already resistant to phage T1; in the presence of phage T1, only these cells survive.

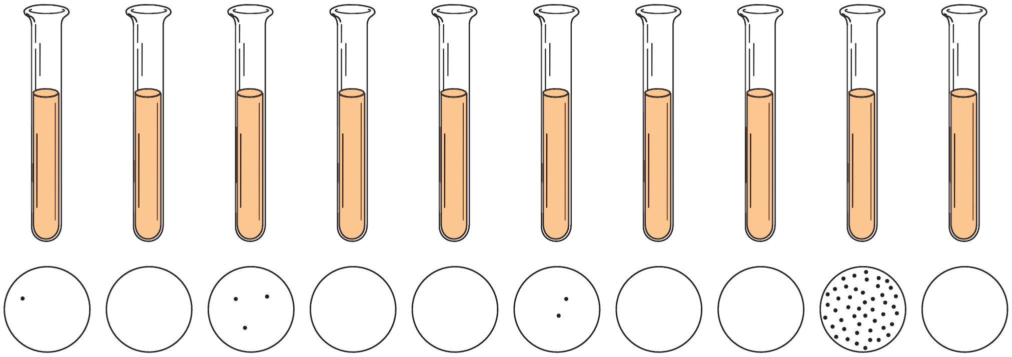
If the presumed rates of physiological induction and mutation are the same, determining which of the two mechanisms is operating is difficult. Luria and Delbrück, however, developed a means of distinguishing between these mechanisms. They reasoned as follows: If T1 resistance was physiologically induced, the relative frequency of resistant *E. coli* cells in a culture of the normal (Ton$^s$) strain should be a constant, independent of the number of cells in the culture or the length of time that the culture has been growing. If resistance was due to random mutation, the frequency of mutant (Ton$^r$) cells would depend on when the mutations occurred. In other words, the appearance of a mutant cell would be a random event. If a mutation occurs early in the growth of the culture, then many cells descend from the mutant cell, and therefore many resistant colonies develop. If the mutation does not occur until late in the growth of the culture, then the subsequent number of mutant cells is few. Thus, if the mutation hypothesis is correct, there should be considerable fluctuation from culture to culture in the number of resistant cells present (fig. 12.1).

#### Results of the Fluctuation Test

To distinguish between these hypotheses, Luria and Delbrück developed what is known as the **fluctuation test.** They counted the mutants both in small ("individual") cultures and in subsamples from a single large ("bulk") culture. All subsamples from a bulk culture should have the same number of resistant cells, differing only because of



(a) Physiological induction



(b) Random mutation

**Figure 12.1** Occurrence of *E. coli* Ton$^r$ colonies in Ton$^s$ cultures. Ten cultures of *E. coli* cells were grown from a standard inoculum in separate test tubes in the absence of phage T1, then spread on petri plates in the presence of phage T1. The resistant cells grow into colonies on the plates. We expect a uniform distribution of resistant cells if the physiological induction hypothesis is correct (*a*) or a great fluctuation in the number of resistant cells if the random mutation hypothesis is correct (*b*).

random sampling error. If, however, mutation occurs, the number of resistant cells among the individual cultures should vary considerably from culture to culture; the number would be related to the time that the mutation occurred during the growth of each culture. If mutation arose early, there would be many resistant cells. If it arose late, there would be relatively few resistant cells. Under physiological induction, the distribution of resistant colonies should not differ between the individual and bulk cultures.

Luria and Delbrück inoculated twenty individual cultures and one bulk culture with *E. coli* cells and incubated them in the absence of phage T1. Each individual culture was then spread out on a petri plate containing a very high concentration of T1 phages; ten subsamples from the bulk culture were plated in the same way. We can see from the results (table 12.1) that there was minimal variation in the number of resistant cells among the bulk culture subsamples but a very large amount of variation, as predicted for random mutation, among the individual cultures.

If bacteria have "normal" genetic systems that undergo mutation, bacteria could then be used, along with higher organisms, to answer genetic questions. As we have pointed out, the modern era of molecular genetics began with the use of prokaryotic and viral systems in genetic research. In the next section, we turn our attention to several basic questions about the gene, questions whose answers were found in several instances only because prokaryotic systems were available.

## Genetic Fine Structure

How do we determine the relationship among several mutations that cause the same phenotypic change? What are the smallest units of DNA capable of mutation and recombination? Are the gene and its protein product colinear? The answers to the latter two questions are important from a historical perspective. The answer to the first question is relevant to our current understanding of genetics.

**Table 12.1** **Results from the Luria and Delbrück Fluctuation Test**

| Individual Cultures* | | Samples from Bulk Culture* | |
|---|---|---|---|
| Culture Number | Ton$^r$ Colonies Found | Sample Number | Ton$^r$ Colonies Found |
| 1 | 1 | 1 | 14 |
| 2 | 0 | 2 | 15 |
| 3 | 3 | 3 | 13 |
| 4 | 0 | 4 | 21 |
| 5 | 0 | 5 | 15 |
| 6 | 5 | 6 | 14 |
| 7 | 0 | 7 | 26 |
| 8 | 5 | 8 | 16 |
| 9 | 0 | 9 | 20 |
| 10 | 6 | 10 | 13 |
| 11 | 107 | | |
| 12 | 0 | | |
| 13 | 0 | | |
| 14 | 0 | | |
| 15 | 1 | | |
| 16 | 0 | | |
| 17 | 0 | | |
| 18 | 64 | | |
| 19 | 0 | | |
| 20 | 35 | | |
| Mean ($\bar{n}$) | 11.4 | | 16.7 |
| Standard deviation | 27.4 | | 4.3 |

*Source:* From E. Luria and M. Delbrück, *Genetics,* 28: 491. Copyright © 1943 Genetics Society of America.

* Each culture and sample was 0.2 ml and contained about $2 \times 10^7$ *E. coli* cells.

### *Complementation*

If two recessive mutations arise independently and both have the same phenotype, how do we know whether they are both mutations of the same gene? That is, how do we know whether they are alleles? To answer this question, we must construct a heterozygote and determine the **complementation** between the two mutations. A heterozygote with two mutations of the same gene will produce only mutant messenger RNAs, which result in mutant enzymes (fig. 12.2*a*). If, however, the mutations are not allelic, the gamete from the $a_1$ parent will also contain an $a_2^+$ allele, and the gamete from the $a_2$ parent will also contain the $a_1^+$ allele (fig. 12.2*b*). If the two mutant genes are truly alleles, then the phenotype of the heterozygote should be mutant. If, however, the two mutant genes are nonallelic, then the $a_1$ mutant will have contributed the wild-type allele at the $A_2$ locus, and the $a_2$ mutant will have contributed the wild-type allele at the $A_1$ locus to the heterozygote. Thus, the two mutations will complement each other and produce the wild-type. Mutations that fail to complement each other are termed **functional alleles.** The test for defining alleles strictly on this basis of functionality is termed the ***cis-trans*** **complementation test.**

There are two different configurations in which a heterozygous double mutant of functional alleles can form (fig. 12.3). In the *cis-trans* complementation test, only the *trans* configuration is used to determine whether the two mutations were allelic. In reality, the *cis* configuration is not tested; it is the conceptual control, in which wild-type activity (with recessive mutations) is always expected. The test is thus sometimes simply called a *trans* test. Functional alleles produce a wild-type phenotype in the *cis* configuration but a mutant phenotype in the *trans* configuration. This difference in phenotypes is called a *cis-trans* position effect.

From the terms *cis* and *trans,* Seymour Benzer coined the term **cistron** for the smallest genetic unit (length of genetic material) that exhibits a *cis-trans* position effect. We thus have a new word for the gene, one in which function is more explicit. We have, in essence, refined Beadle and Tatum's one-gene-one-enzyme hypothesis to a more accurate one-cistron-one-polypeptide concept. The cistron is the smallest unit that codes for a messenger RNA that is then translated into a single polypeptide or expressed directly (transfer RNA or ribosomal RNA).

From functional alleles, we can go one step further in recombinational analysis by determining whether two allelic mutations occur at exactly the same place in the cistron. In other words, when two mutations prove to be functional alleles, are they also **structural alleles?** The methods used to analyze complementation can be used here also. Crosses are carried out to form a mutant heterozygote (*trans* configuration) whose offspring are then tested for recombination between the two mutational sites. If no recombination occurs, then the two alleles probably contain the same

Seymour Benzer (1921– ).
(Courtesy of Dr. Seymour Benzer, 1970.)

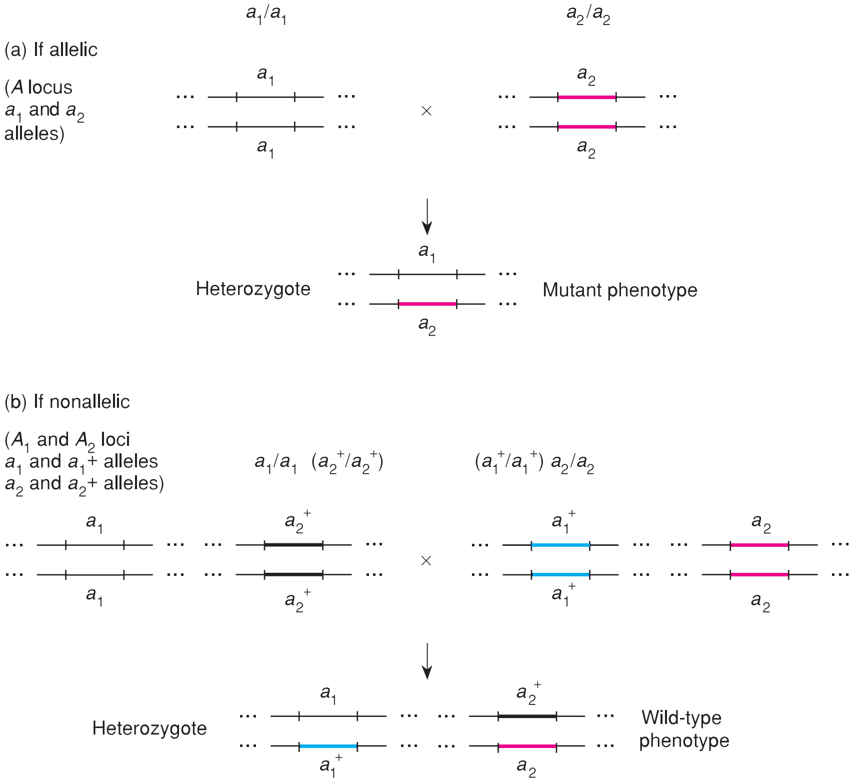

(a) If allelic
(A locus a1 and a2 alleles)
a1/a1
a2/a2
a1
a1
a2
a2
×
Heterozygote
Mutant phenotype
(b) If nonallelic
(A1 and A2 loci a1 and a1+ alleles a2 and a2+ alleles)
a1/a1 (a2+/a2+)
(a1+/a1+) a2/a2
a1
a2+
a1+
a2
Heterozygote
Wild-type phenotype

**Figure 12.2** The complementation test defines allelism. Are two mutations ($a_1$, $a_2$) allelic if they affect the same trait? To find out, mutant homozygotes are crossed to form a heterozygote. (*a*) If the mutations are allelic, then both copies of the gene in the heterozygote are mutant, resulting in the mutant phenotype. (*b*) If the mutations are nonallelic, then there is a wild-type allele of each gene present in the heterozygote, resulting in the wild-type phenotype. (The two loci need not be on the same chromosome.)

structural change (involving the same base pairs) and are thus structural alleles. If a small amount of recombination occurs that generates wild-type offspring, then the two alleles are not mutations at the same point (fig. 12.4). Alleles that were functional but not structural were first termed **pseudoalleles** because it was believed that loci were made up of subloci. Fine-structure analysis led to the understanding that a locus is a length of genetic material divisible by recombination rather than a "bead on a string."

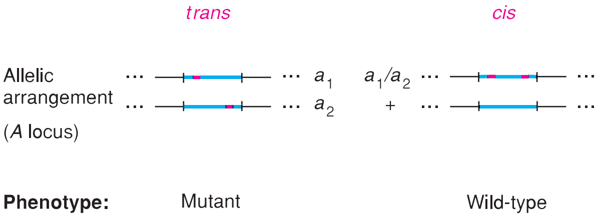

trans
cis
Allelic arrangement
(A locus)
a1
a2
a1/a2
+
Phenotype:
Mutant
Wild-type

**Figure 12.3** A heterozygote of two recessive mutations can have either the *trans* or *cis* arrangement. In the *trans* position, functional alleles produce a mutant phenotype. (*Red marks* represent mutant lesions.) In the *cis* position, functional alleles produce a wild-type phenotype. The *cis-trans* position effect thus reveals functional alleles.

Eye-color mutants of *Drosophila melanogaster* can be studied by complementational analysis. The white-eye locus has a series of alleles producing varying shades of red. This locus is sex linked, at about map position 3.0 on the X chromosome. (Several other eye-color loci on the X chromosome are not relevant to this cross—e.g., prune and ruby.) If an apricot-eyed female is mated with a white-eyed male, the female offspring are all heterozygous and have mutant light-colored eyes (fig. 12.5). Thus, apricot and white are functional alleles: they do not complement (table 12.2). To determine whether apricot and white are structural alleles, light-eyed females are crossed with white-eyed males, and the offspring are observed for the presence of wild-type or light-eyed males. Though their rate of appearance is less than 0.001%, this is significantly above the background mutation rate. The conclusion is that apricot and white are functional, but not structural, alleles.

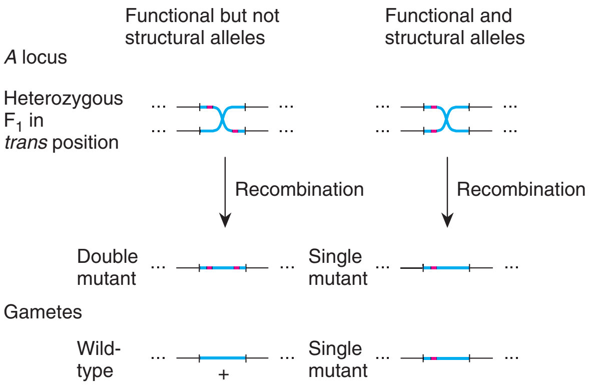

Functional but not structural alleles
Functional and structural alleles
A locus
Heterozygous F1 in trans position
Recombination
Recombination
Double mutant
Single mutant
Gametes
Wild-type
+
Single mutant

**Figure 12.4** Functional alleles may or may not be structurally allelic. (*Red marks* represent mutant sites.) Functional alleles that are not also structural alleles can recombine between the mutant sites, resulting in occasional wild-type (and double mutant) offspring. Structural alleles (which are also always functional alleles) are defective at the same base pairs and cannot form either wild-type or double mutant offspring by recombination.

## Fine-Structure Mapping

After Beadle and Tatum established in 1941 that a gene controls the production of an enzyme that then controls a step in a biochemical pathway, Benzer used analytical techniques to dissect the fine structure of the gene. Fine-structure mapping means examining the size and number of sites within a gene that are capable of mutation and recombination. In the late 1950s, when biochemical techniques were not yet available for DNA sequencing, Benzer used classical recombinational and mutational techniques with bacterial viruses to provide reasonable estimates on the details of fine structure and to give insight into the nature of the gene. He coined the terms **muton** for the smallest mutable site and **recon** for the smallest unit of recombination. It is now known that both muton and recon are a single base pair.

Before Benzer's work, genes were thought of as beads on a string. The very low rate of recombination between

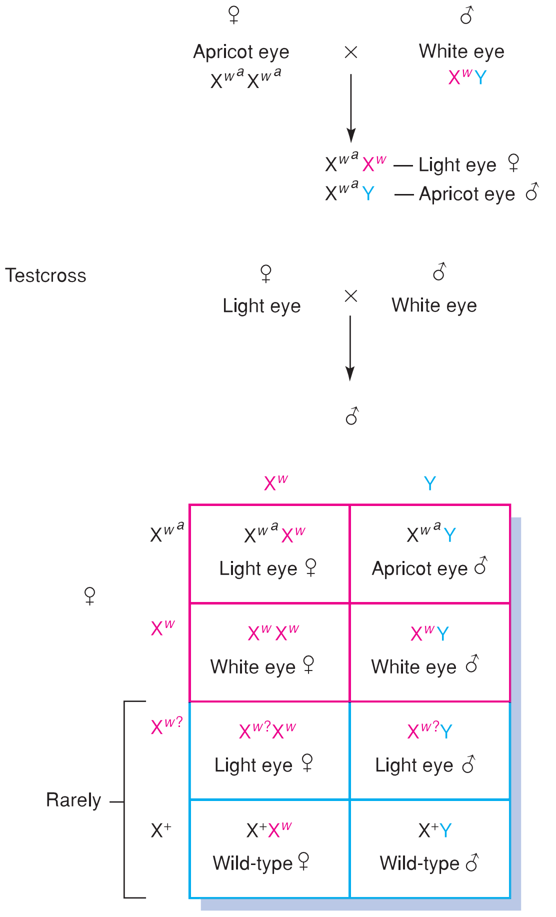

♀
Apricot eye
$X^{w^a}X^{w^a}$
×
♂
White eye
$X^wY$
$X^{w^a}X^w$ — Light eye ♀
$X^{w^a}Y$ — Apricot eye ♂
Testcross
♀
Light eye
×
♂
White eye
♂
$X^w$
Y
♀
$X^{w^a}$
$X^{w^a}X^w$
Light eye ♀
$X^{w^a}Y$
Apricot eye ♂
$X^w$
$X^wX^w$
White eye ♀
$X^wY$
White eye ♂
Rarely
$X^{w?}$
$X^{w?}X^w$
Light eye ♀
$X^{w?}Y$
Light eye ♂
$X^+$
$X^+X^w$
Wild-type ♀
$X^+Y$
Wild-type ♂

**Figure 12.5** Crosses demonstrating that apricot and white eyes are functional, but not structural, alleles in *Drosophila.* Light-eyed females are heterozygous for both alleles. When testcrossed, they produce occasional offspring that are wild-type ($X^+$ allele) or light-eyed ($X^{w?}$ allele). This indicates a crossover between the two mutant sites (white and apricot) in the heterozygous females, producing, reciprocally, an allele with both mutational sites and the wild-type.

**Table 12.2** Complementation Matrix of X-Linked *Drosophila* Eye-Color Mutants

| | white | prune | apricot | buff | cherry | eosin | ruby |
|---|---|---|---|---|---|---|---|
| white (*w*) | – | + | – | – | – | – | + |
| prune (*pn*) | | – | + | + | + | + | + |
| apricot ($w^a$) | | | – | – | – | – | + |
| buff ($w^{bf}$) | | | | – | – | – | + |
| cherry ($w^{ch}$) | | | | | – | – | + |
| eosin ($w^e$) | | | | | | – | + |
| ruby (*rb*) | | | | | | | – |

*Note:* Plus sign indicates that female offspring are wild-type; minus sign indicates that they are mutant.

sites within a gene hampered the analysis of mutational sites within a gene by means of recombination. If two mutant genes are functional alleles (involving different sites on the same gene), a distinct probability exists that we will get both mutant sites (and both wild-type sites) on the same chromosome by recombination (see fig. 12.4); but, in view of the very short distances within a gene, this probability is very low. Although it certainly seemed desirable to map sites within the gene, the problem of finding an organism that would allow fine-structure analysis remained until Benzer decided to use phage T4.

**r II Screening Techniques.** Benzer used the T4 bacteriophage because of the growth potential of phages, in which a generation takes about an hour and the increase in numbers per generation is about a hundredfold. Actually, any prokaryote or virus should suffice, but Benzer made use of other unique screening properties of the phage that made it possible to recognize one particular mutant in about a billion phages. Benzer used *r*II mutants of T4. These mutants produce large, smooth-edged plaques on *E. coli,* whereas the wild-type produces smaller plaques whose edges are not as smooth (see fig. 7.7).

The screening system that Benzer employed made use of the fact that *r*II mutants do not grow on *E. coli* strain K12, whereas the wild-type can. The normal host strain, *E. coli* B, allows growth of both the wild-type and *r*II mutants. Thus, various mutants can be crossed by mixed infection of *E. coli* B cells, and Benzer could screen for wild-type recombinants by plating the resultant progeny phages on *E. coli* K12 (fig. 12.6), on which only a wild-type recombinant produces a plaque. It is possible to detect about one recombinant in a billion phages, all in an afternoon's work. This ability to detect recombinants occurring at such a low level of frequency allowed Benzer to see recombinational events occurring very close together on the DNA, events that would normally occur at a frequency too low to detect in fruit flies or corn.

Benzer sought to map the number of sites subject to recombination and mutation within the *r*II region of T4. He began by isolating independently derived *r*II mutants and crossing them among themselves. The first thing he found was that the *r*II region was composed of two cistrons; almost all of the mutations belonged to one of two **complementation groups.** The *A*-cistron mutations would not complement each other but would complement the mutations of the *B* cistron. The exceptions were mutations that seemed to belong to both cistrons. These mutations were soon found to be deletions in which part of each cistron was missing (table 12.3).

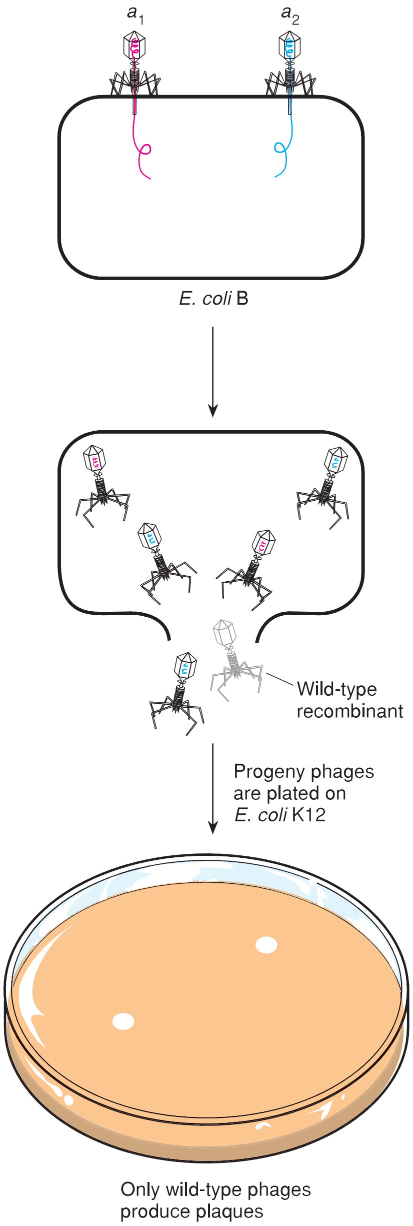

$a_1$
$a_2$
E. coli B
Wild-type recombinant
Progeny phages are plated on E. coli K12
Only wild-type phages produce plaques

**Figure 12.6** Using *E. coli* K12 and B strains to screen for recombination at the *r*II locus of phage T4. Two *r*II mutants are crossed by infecting the same B-strain bacteria with both phages. The offspring are plated on a lawn of K12 bacteria in which only wild-type phages can grow. The technique thus selects only wild-type recombinants.

**Deletion Mapping.** As the number of independently isolated mutations of the *A* and *B* cistrons increased, it became obvious that to make every possible pairwise cross would entail millions of crosses. To overcome this problem, Benzer isolated mutants that had partial or complete deletions of each cistron. Deletion mutations were easy to discover because they acted like structural alleles to alleles that were not themselves structurally allelic. In other words, if mutations *a, b,* and *c* are functional—but not structural—alleles of each other, and mutation *d* is a structural allele to *a, b,* and *c,* then *d* must contain a deletion of the bases mutated in *a, b,* and *c.* Once a sequence of deletion mutations covering the *A* and *B* cistrons was isolated, a minimal number of crosses was required to localize a new mutation to a portion of one of the cistrons. A second series of smaller deletions within each region was then isolated, further localizing the mutation (fig. 12.7).

Next, each new mutant was crossed with each of the other mutants isolated in its subregion to localize the relative position of the new mutation. If the mutation was structurally allelic to a previously isolated mutation, it was scored as an independent isolation of the same mutation. If it was not a structural allele to any of the known mutations of the subregion, it was added as a new mutation point. The exact position of each new mutation within the region was determined by the relative frequency of recombination between it and the known mutations of this region (see chapter 7). Benzer eventually isolated about 350 mutations from eighty different subregions defined by deletion mutations. An abbreviated map is shown in figure 12.8.

What conclusions did Benzer draw from his work? First, he concluded that since all of the mutations in both *r*II cistrons can be ordered in a linear fashion, the original Watson-Crick model of DNA as a linear molecule was correct. Second, he concluded that reasonable inroads had been made toward saturating the map, localizing at least one mutation at every mutable site. Benzer reasoned that since many sites were represented by only one mutation, some sites must occur that were represented by zero mutations (i.e., not yet represented by a mutation). Since he had mapped about 350 sites, he calculated that there

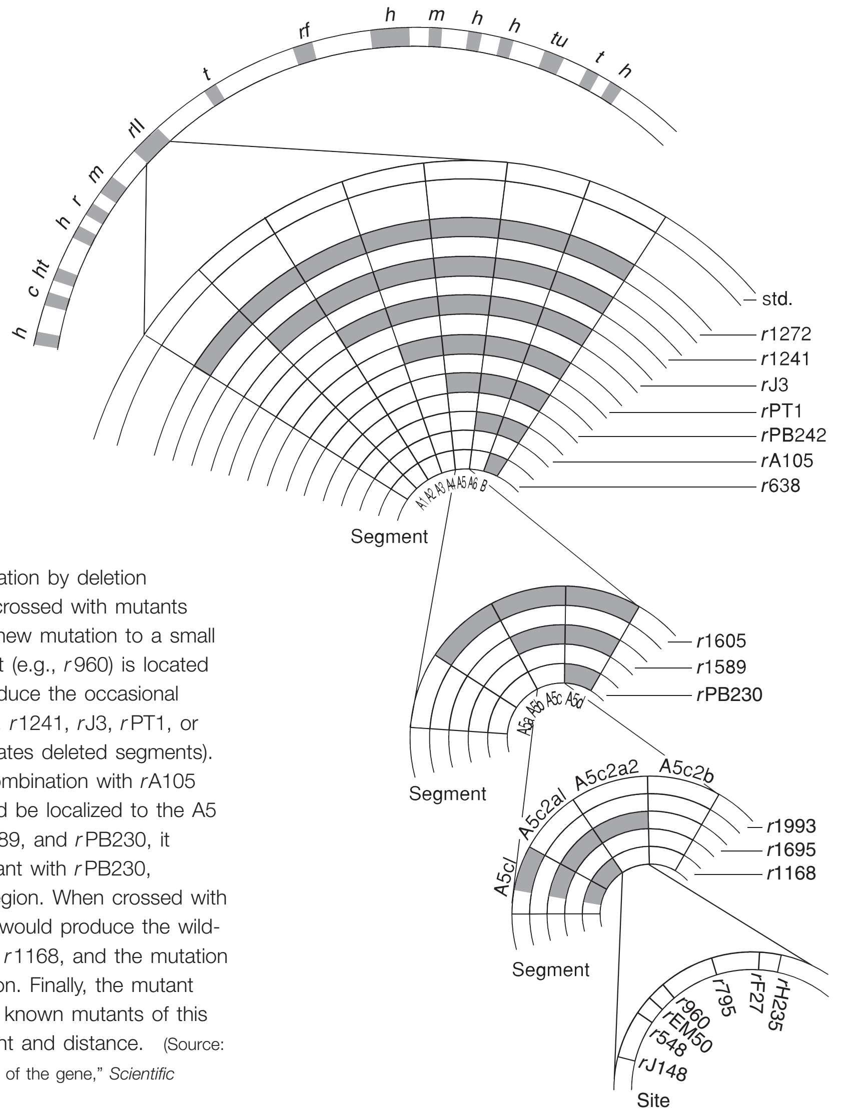


**Figure 12.7** Localization of an *r*II mutation by deletion mapping. Newly isolated mutants are crossed with mutants with selected deletions to localize the new mutation to a small region of the cistron. If the new mutant (e.g., *r*960) is located in the A5c2a2 region, it would not produce the occasional wild-type by recombination with *r*1272, *r*1241, *r*J3, *r*PT1, or *r*PB242 (the *solid* part of the bar indicates deleted segments). It would produce the wild-type by recombination with *r*A105 and *r*638, and thus the mutation would be localized to the A5 region. When crossed with *r*1605, *r*1589, and *r*PB230, it would produce only the rare recombinant with *r*PB230, indicating the mutation is in the A5c region. When crossed with *r*1993, *r*1695, and *r*1168, the mutant would produce the wild-type by recombination with *r*1993 and *r*1168, and the mutation would be localized to the A5c2a2 region. Finally, the mutant would be crossed pairwise with all the known mutants of this region to determine relative arrangement and distance. (Source: Data from Seymour Benzer, "The fine structure of the gene," *Scientific American,* 206: 70–84, January 1962.)

**Table 12.3** **Complementation Matrix of Ten *r*II Mutants**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | – | – | + | – | – | – | + | – | – | – |
| 2 | | – | + | – | – | – | + | – | – | – |
| 3 | | | – | – | + | + | – | + | – | + |
| 4 | | | | – | – | – | – | – | – | – |
| 5 | | | | | – | – | + | – | – | – |
| 6 | | | | | | – | + | – | – | – |
| 7 | | | | | | | – | + | – | + |
| 8 | | | | | | | | – | – | – |
| 9 | | | | | | | | | – | – |
| 10 | | | | | | | | | | – |

*Note:* Plus sign indicates complementation; minus sign indicates no complementation. The two cistrons are arbitrarily designated *A* and *B*. Mutants 4 and 9 must be deletions that cover parts of both cistrons. Alleles: *A* cistron: 1, 2, 4, 5, 6, 8, 9, 10; *B* cistron: 3, 4, 7, 9.

were at least another 100 sites still undetected by mutation. We now know that 450 sites is an underestimate. However, since the protein products of these cistrons were not isolated, there were no independent estimates of the number of nucleotides in these cistrons (number of amino acids times three nucleotides per codon). Thus, although Benzer had not saturated the map with mutations, he certainly had made respectable progress in dissecting the gene and demonstrating that it was not an indivisible unit, a "bead on a string."

**Hot Spots.** Benzer also looked into the lack of uniformity in the occurrence of mutations (note two major "hot spots" at B4 and A6c of fig. 12.8). Presuming that all base pairs are either AT or GC, this lack of uniformity was unexpected. Benzer suggested that spontaneous mutation is not just a function of the base pair itself, but is affected by the surrounding bases as well. This concept still holds.

To recapitulate, Benzer's work supports the model of the gene as a linear arrangement of DNA whose nucleotides are the smallest units of mutation. The link between any adjacent nucleotides can break in the recombinational process. The smallest functional unit, determined by a complementation test, is the cistron. Mutagenesis is not uniform throughout the cistron, but may depend on the particular arrangement of bases in a given region.

## *Intra-Allelic Complementation*

Benzer warned that certainty is elusive in the complementation test because sometimes two mutations of the same functional unit (cistron) can result in partial activity. The

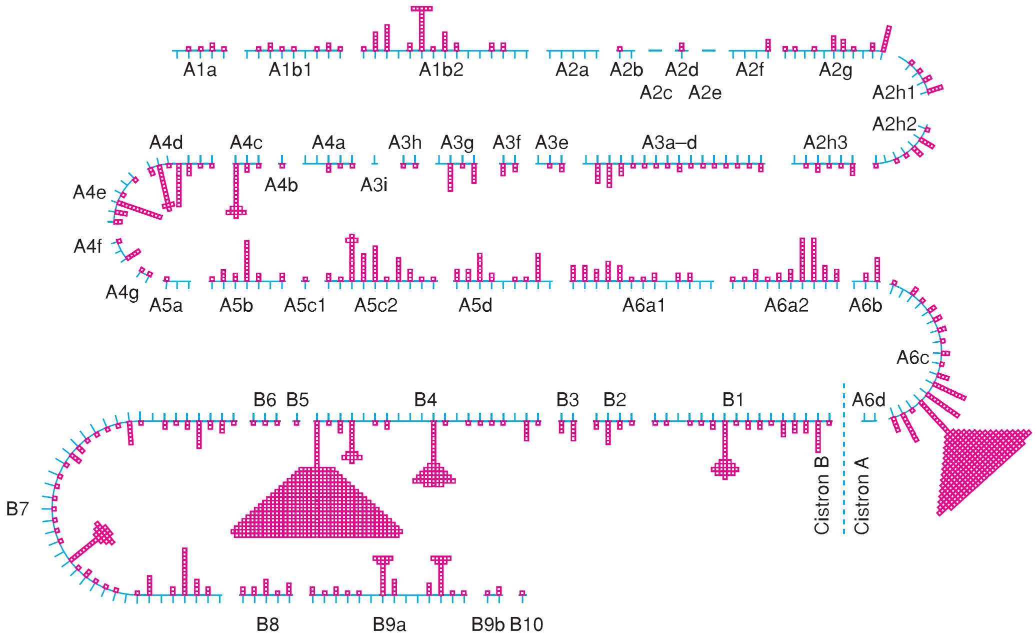


**Figure 12.8** Abbreviated map of spontaneous mutations of the *A* and *B* cistrons of the *r*II region of T4. Each square represents one independently isolated mutation. Note the "hot spots" at A6c and B4. (From Seymour Benzer, "On the topography of the genetic fine structure", *Proceedings of the National Academy of Sciences USA* 47:403–15, 1961. Reprinted by permission.)

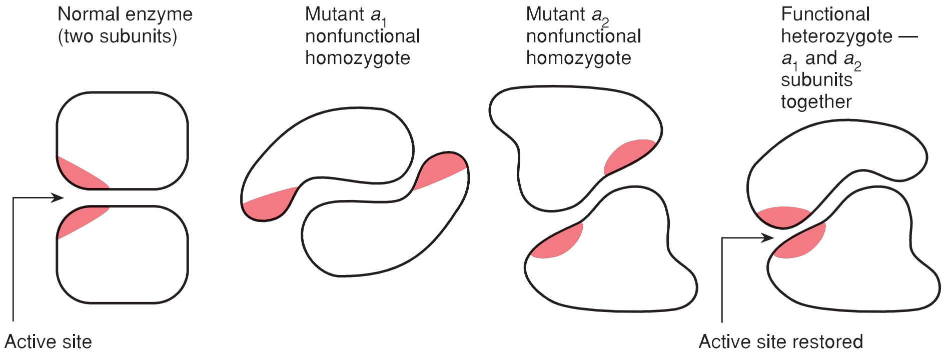

Normal enzyme (two subunits)
Mutant $a_1$ nonfunctional homozygote
Mutant $a_2$ nonfunctional homozygote
Functional heterozygote — $a_1$ and $a_2$ subunits together
Active site
Active site restored

**Figure 12.9** Intra-allelic complementation. With certain mutations, it is possible to get enzymatic activity in a heterozygote for two nonfunctional alleles, if the two polypeptides form a functional enzyme. (Active site is shown in *color.*)

problem can be traced to the interactions of subunits at the polypeptide level. Some proteins are made up of subunits, and it is possible that certain mutant combinations produce subunits that interact to restore the enzymatic function of the protein (fig. 12.9). This phenomenon is known as **intra-allelic complementation.** With this in mind, geneticists routinely use the complementation test to determine functional relationships among mutations.

## Colinearity

Next we look at the colinearity of the gene and the polypeptide. Benzer's work established that the gene was a linear entity, as Watson and Crick had proposed. However, Benzer could not demonstrate the colinearity of the gene and its protein product. To do this, it is necessary to show that for every mutational change in the DNA, a corresponding change takes place in the protein product of the gene. Colinearity would be established by showing that nucleotide and amino acid changes occurred in a linear fashion and in the same order in the protein and in the cistron.

Ideally, Benzer himself might have solved the colinearity issue. He was halfway there, with his 350 or so isolated mutations of phage T4. However, Benzer did not have a protein product to analyze; no mutant protein had been isolated from *r*II mutants. In the midst of competition to find just the right system, Charles Yanofsky of Stanford University and his colleagues emerged in the mid-1960s with the required proof, showing that the order of a polypeptide's amino acids corresponded to the nucleotide sequence in the gene that specified it. Yanofsky's success rested with his choice of an amenable system, one using the enzymes from a biochemical pathway.

Yanofsky did his research on the tryptophan biosynthetic pathway in *E. coli.* The last enzyme in the pathway, tryptophan synthetase, catalyzes the reaction of indole-3-glycerol-phosphate plus serine to tryptophan and 3-phosphoglyceraldehyde. The enzyme itself is made of four subunits specified by two separate cistrons, with each polypeptide present twice.

Yanofsky and his colleagues concentrated on the *A* subunit. They mapped *A*-cistron mutations with transduction (see chapter 7) using the transducing phage P1. They first tested each new mutant against a series of deletion mutants to establish the region where the mutation was. Then they crossed mutants for a particular region among themselves to establish relative positions and distances.

The protein products of the bacterial genes were isolated using electrophoresis and chromatography to establish the fingerprint patterns of the proteins (see chapter 11). Assuming a single mutation, a comparison of the mutant and the wild-type fingerprints would show a difference of just one polypeptide spot (fig. 12.10), avoiding the need to sequence the entire protein. The mutant amino acid was

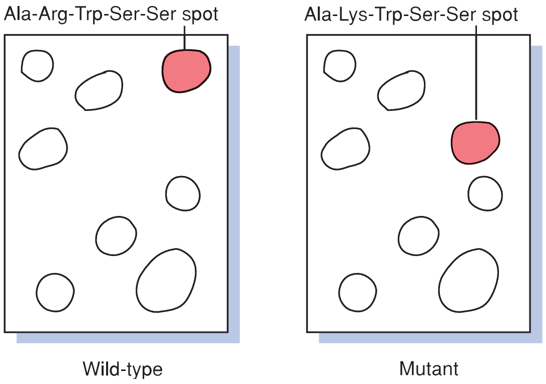


**Figure 12.10** Difference in "fingerprints" between mutant and wild-type polypeptide digests. The single spot that differs in the mutant can be isolated and sequenced, eliminating the need to sequence the whole protein.

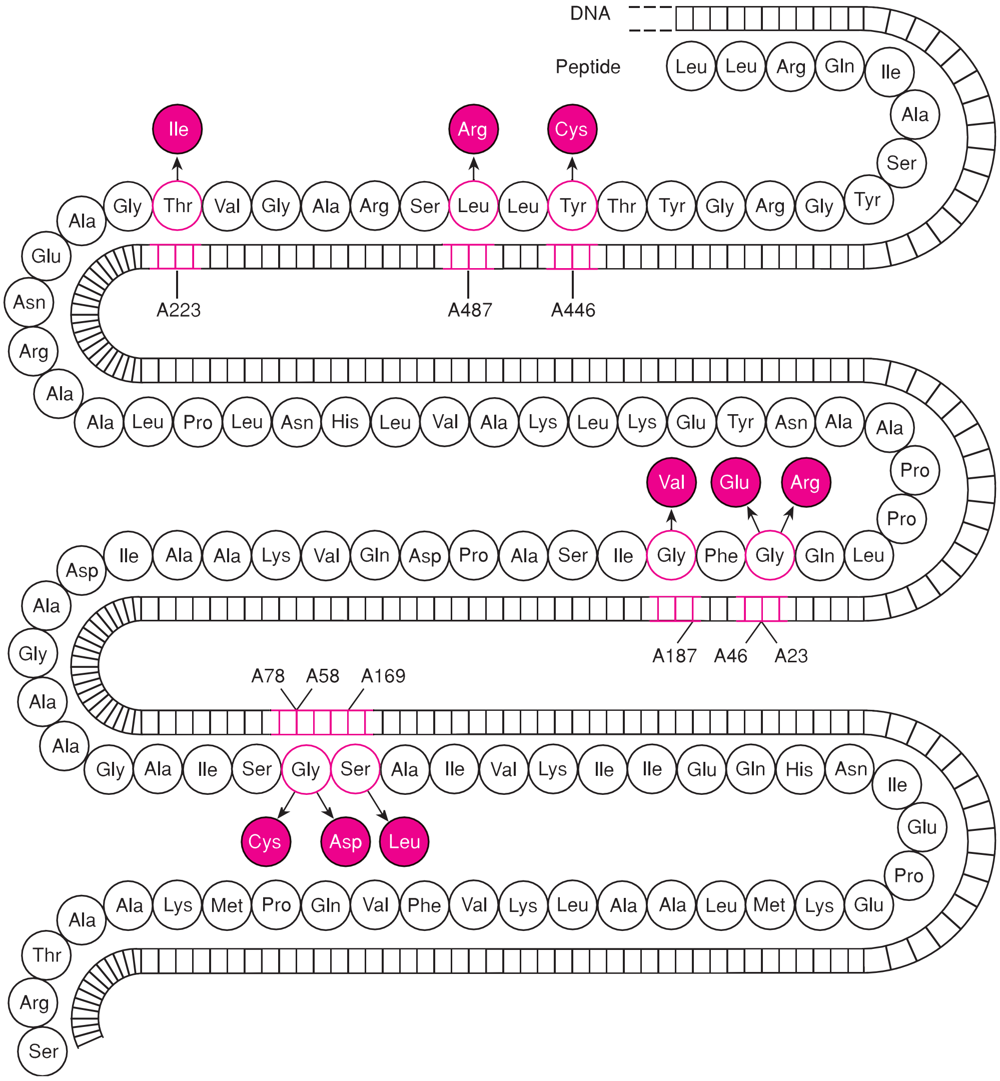


**Figure 12.11** Amino acid sequence of the carboxyl terminal end of the tryptophan synthetase *A* protein and its DNA. Mutations are shown on the DNA (e.g., A446), as are the changed amino acids of those mutations (in *color*). DNA and protein changes are colinear.

identified by analysis of just this one spot. Figure 12.11 shows the details of nucleotide and amino acid changes for nine of the mutations in this 267-amino acid protein.

We can see from this figure that nine mutations in the linear *A* cistron of tryptophan synthetase are colinear with nine amino acid changes in the protein itself. In two cases, two mutations mapped so close as to be almost indistinguishable. In both cases, the two mutations proved to be in the same codon: the same amino acid position was altered in each (A23–A46, A58–A78). Thus, exactly as predicted and expected, colinearity exists between gene and protein. Brenner and his colleagues, using head-protein mutants of phage T4, independently confirmed this work at the same time.

## Spontaneous Versus Induced Mutation

H. J. Muller won the Nobel Prize for demonstrating that X rays can cause mutations. This work was published in 1927 in a paper entitled "Artificial Transmutation of the Gene." At about the same time, L. J. Stadler induced mutations in barley with X rays. The basic impetus for their work was the fact that mutations occur so infrequently that genetic research was hampered by the inability to

Hermann J. Muller (1890–1967). (Courtesy of National Academy of Sciences.)

Lewis J. Stadler (1896–1954). (*Genetics,* 41, 1956: frontispiece.)

obtain mutants. Muller exposed flies to varying doses of X rays and then observed their progeny. He came to several conclusions. First, X rays greatly increased the occurrence of mutations. Second, the inheritance patterns of X-ray-induced mutations and the resulting phenotypes of organisms were similar to those that resulted from natural, or "spontaneous," mutations.

## Mutation Rates

The **mutation rate** is the number of mutations that arise per cell division in bacteria and single-celled organisms, or the number of mutations that arise per gamete in higher organisms. Mutation rates vary tremendously depending upon the length of genetic material, the kind of mutation, and other factors. Luria and Delbrück, for example, found that in *E. coli* the mutation rate per cell division of $Ton^s$ to $Ton^r$ was $3 \times 10^{-8}$, whereas the mutation rate of the wild-type to the histidine-requiring phenotype ($His^+$ to $His^-$) was $2 \times 10^{-6}$. The rate of **reversion** (return of the mutant to the wild-type) was $7.5 \times 10^{-9}$. The mutation and reversion rates differ because many different mutations can cause the His phenotype, whereas reversion requires specific, and hence less probable, changes to correct the His phenotype back to the wild-type. The lethal mutation rate in *Drosophila* is about $1 \times 10^{-2}$ per gamete for the total genome. This number is relatively large because, as with His, many different mutations produce the same phenotype (lethality, in this case).

## Point Mutations

The mutations of primary concern in this chapter are **point mutations,** which consist of single changes in the nucleotide sequence. (In chapter 8 we discussed chromosomal mutations, changes in the number and visible structures of chromosomes.) If the change is a replacement of some kind, then a new codon is created. In many cases, this new codon, upon translation, results in a new amino acid. As discussed in chapter 11, one of the outcomes of redundancy in the genetic code is partial protection of the cell from the effects of mutation; common amino acids have the most codons, similar amino acids have similar codons, and the wobble position of the codon is the least important position in translation. However, when base changes result in new amino acids, new proteins appear. These new proteins can alter the morphology or physiology of the organism and result in phenotypic novelty or lethality.

### Frameshift Mutation

A point mutation may consist of replacement, addition, or deletion of a base (fig. 12.12). Point mutations that add

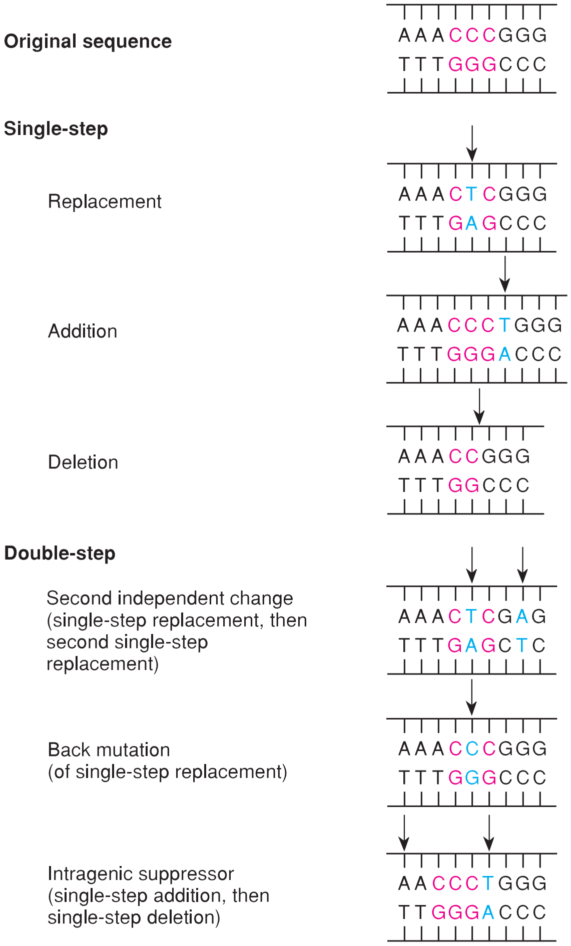

Original sequence
AAACCCGGG
TTTGGGCCC
Single-step
Replacement
AAACTCGGG
TTTGAGCCC
Addition
AAACCCTGGG
TTTGGGACCC
Deletion
AAACCGGG
TTTGGCCC
Double-step
Second independent change (single-step replacement, then second single-step replacement)
AAACTCGAG
TTTGAGCTC
Back mutation (of single-step replacement)
AAACCCGGG
TTTGGGCCC
Intragenic suppressor (single-step addition, then single-step deletion)
AACCCTGGG
TTGGGACCC

**Figure 12.12** Types of DNA point mutations. Single-step changes are replacements, additions, or deletions. A second point mutation in the same gene can result either in a double mutation, reversion to the original, or intragenic suppression. In this case, intragenic suppression is illustrated by the addition of one base followed by the nearby deletion of a different base.

or subtract a base are, potentially, the most devastating in their effects on the cell or organism because they change the reading frame of a gene from the site of mutation onward (fig. 12.13). A frameshift mutation causes two problems. First, all the codons from the frameshift on will be different and thus yield (most probably) a useless protein. Second, stop-signal information will be misread. One of the new codons may be a nonsense codon, which causes translation to stop prematurely. Or, if the translation apparatus reaches the original nonsense codon, it is no longer recognized as such because it is in a different reading frame, and therefore, the translation process continues beyond the end of the gene.

### *Back Mutation and Suppression*

A second point mutation in the same gene can have one of three possible effects (see fig. 12.12). First, the mutation can result in either another mutant codon or in one codon that has experienced two changes. Second, if the change is at the same site, the original sequence can be returned, an effect known as **back mutation:** the gene then becomes a revertant, with its original function restored. Third, **intragenic suppression** can take place. Intragenic suppression occurs when a second mutation in the same gene masks the occurrence of the original mutation without actually restoring the original sequence. The new sequence is a double mutation that appears to have the original (unmutated) phenotype. In figure 12.12, a T addition is followed by an A deletion that substitutes the AACCCT sequence for the original AAACCC. These sequences, when transcribed (UUGGGA, UUUGGG), are codons for leucine-glycine and phenylalanine-glycine, respectively. Intragenic suppression occurs whether the new codons are for different amino acids or the same amino acids, as long as the phenotype of the organism is reverted approximately to the original. Suppressed mutations can be distinguished from true back mutations either by subtle differences in phenotype, by genetic crosses, by changes in the amino acid sequence of a protein, or by DNA sequencing.

### *Conditional Lethality*

A class of mutants that has been very useful to geneticists is the conditional-lethal mutant, a mutant that is lethal under one set of circumstances but not under another set. **Nutritional-requirement mutants** are good examples (see chapter 7). **Temperature-sensitive mutants** are conditional-lethal mutants that have made it possible for geneticists to work with genes that control vital functions of the cell, such as DNA synthesis. Many temperature-sensitive mutants are completely normal at 25° C but cannot synthesize DNA at 42° C. Presumably, temperature-sensitive mutations result in enzymes with amino acid substitutions that cause protein denaturation to occur at temperatures above normal. Thus, the enzyme has normal function at 25° C, the **permissive temperature,** but is nonfunctional at 42° C, the **restrictive temperature.**

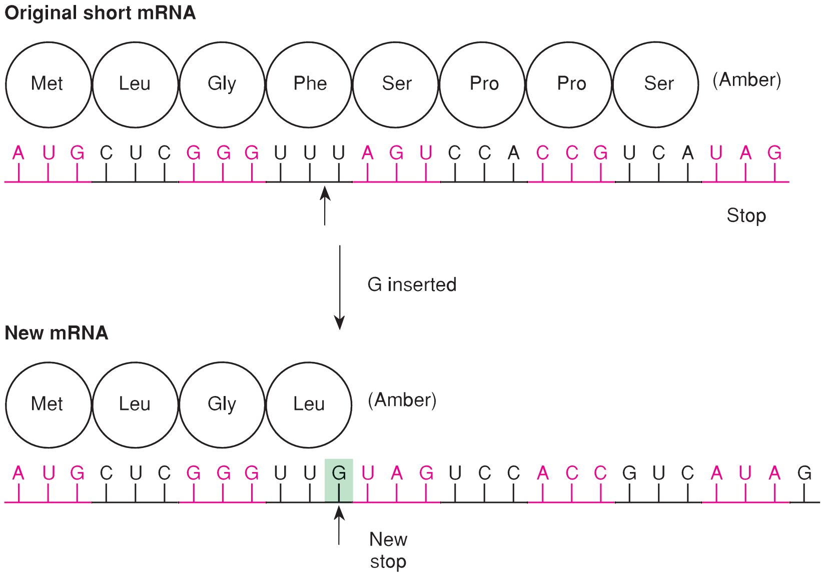

Original short mRNA
Met
Leu
Gly
Phe
Ser
Pro
Pro
Ser
(Amber)
A U G C U C G G G U U U A G U C C A C C G U C A U A G
Stop
G inserted
New mRNA
Met
Leu
Gly
Leu
(Amber)
A U G C U C G G G U U G U A G U C C A C C G U C A U A G
New stop

**Figure 12.13** Possible effects of a frameshift mutation. The insertion of a single base results in the creation of a new stop sequence (*amber*). The result will be premature termination of translation.

The interesting thing about most conditional-lethal mutants of *E. coli* that cannot synthesize DNA at the restrictive temperature is that they have a completely normal DNA polymerase I. From this information, we infer that polymerase I is not the enzyme *E. coli* normally uses for DNA replication. When an organism with a conditional mutation of polymerase I was isolated, it was able to replicate its DNA normally, but unable to repair damage to the DNA. This led to the conclusion that polymerase I is primarily involved in repair rather than replication of DNA. Conditional-lethal mutants thus allow genetic analysis on genes otherwise impossible to study.

## *Spontaneous Mutagenesis*

Watson and Crick originally suggested that mutation could occur spontaneously during DNA replication if pairing errors occurred. If a base of the DNA underwent a proton shift into one of its rare tautomeric forms **(tautomeric shift)** during the replication process, an inappropriate pairing of bases would occur. Normally, adenine and cytosine are in the amino ($NH_2$) form. Their tautomeric shifts are to the imino (NH) form. Similarly, guanine and thymine go from a keto (C=O) form to an enol (COH) form (fig. 12.14). Table 12.4 shows the new

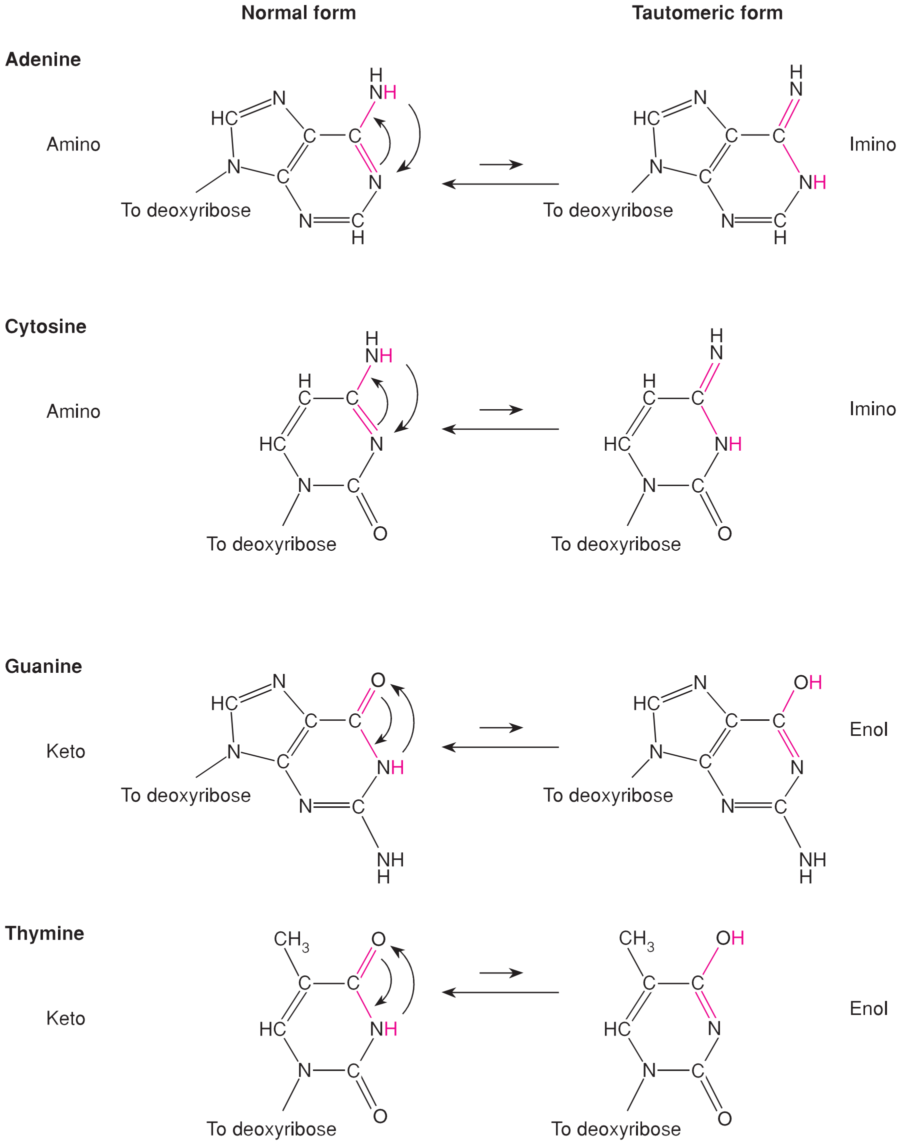


**Figure 12.14** Normal and tautomeric forms of DNA bases. Adenine and cytosine can exist in the amino, or the rare imino, forms; guanine and thymine can exist in the keto, or rare enol, forms.

**Table 12.4** **Pairing Relationships of DNA Bases in the Normal and Tautomeric Forms**

| Base | In Normal State Pairs with | In Tautomeric State Pairs with |
|---|---|---|
| A | T | C |
| T | A | G |
| G | C | T |
| C | G | A |

base pairings that would occur following tautomeric shifts of the DNA bases. Figure 12.15 illustrates the molecular structure of one of these tautomeric pairings.

During DNA replication, a tautomeric shift in either the incoming base (*substrate transition*) or the base already in the strand (*template transition*) results in mispairing. The mispairing will be permanent and result in a new base pair after an additional round of DNA replication. The original strand is unchanged (fig. 12.16).

In the example in figure 12.16, the replacement of one base pair maintains the same purine-pyrimidine relationship: AT is replaced by GC and GC by AT. In both examples, a purine-pyrimidine combination is replaced by a purine-pyrimidine combination. (Or, more specifically, a purine replaces another purine: guanine replaces adenine in the first example and adenine replaces guanine in the second.) The mutation is referred to as a **transition mutation:** a purine (or pyrimidine) replaces another purine (or pyrimidine) through a transitional state involving a tautomeric shift. When a purine replaces a pyrimidine or vice versa, it is referred to as a **transversion mutation.**

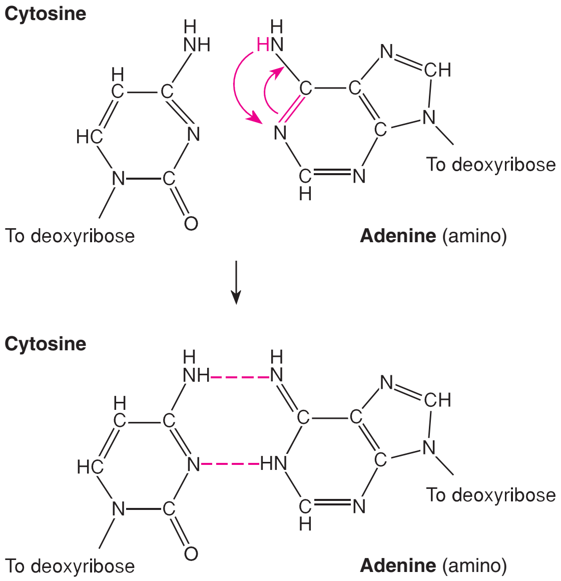
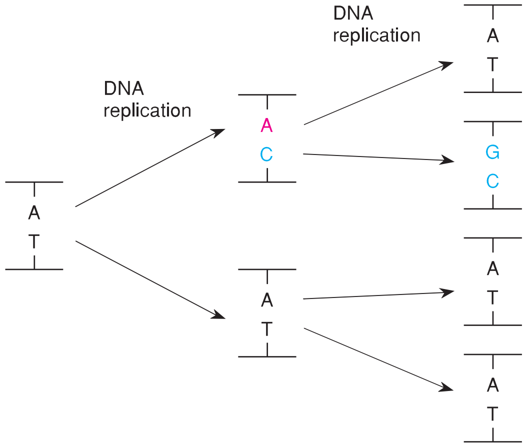
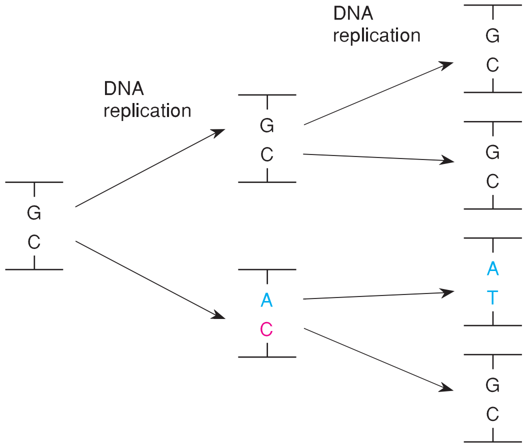

Cytosine
H
NH
H
HN
N
CH
H
C
C
C
C
HC
N
N
C
N
C
N
O
H
To deoxyribose
To deoxyribose
Adenine (amino)
Cytosine
H
NH
H
N
N
CH
H
C
C
C
C
HC
N
HN
N
C
N
C
N
O
H
To deoxyribose
To deoxyribose
Adenine (amino)

**Figure 12.15** Tautomeric forms of adenine. In the common amino form, adenine does not base-pair with cytosine; in the tautomeric imino form, it can.


DNA replication
DNA replication
A
T
A
C
A
T
A
T
G
C
A
T
A
T

Template transition—tautomerization of adenine in the template


DNA replication
DNA replication
G
C
G
C
A
C
G
C
G
C
A
T
G
C

Substrate transition—tautomerization of incoming adenine

**Figure 12.16** Tautomeric shifts result in transition mutations. The tautomerization can occur in the template base or in the substrate base. Tautomeric shifts are shown in *red;* the resulting transition in *blue.* The transition shows up after a second generation of DNA replication.

Transversions may arise by a combination of two events, a tautomerization and a base rotation. (We saw base rotations in the formation of Z DNA in chapter 9.) For example, an AT base pair can be converted to a TA base pair (a transversion) by an intermediate AA pairing (fig. 12.17). Adenine can pair with adenine if one of the bases undergoes a tautomeric shift while the other rotates about its base-sugar (glycosidic) bond (fig. 12.18). The normal configuration of the base is referred to as the *anti* configuration; the rotated form is the *syn* configuration. Since we now believe that as many as 10% of bases may be in the *syn* configuration at any moment, the transversion mutagenesis rate should be about 10% of the transition mutagenesis rate, a value not inconsistent with current information.

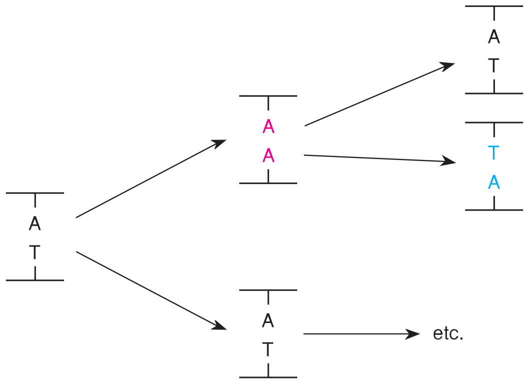


**Figure 12.17** A model for transversion mutagenesis. An AT base pair can be converted to a TA base pair (a transversion) by way of an intermediate AA base pair. One of the *red* bases is in the rare tautomeric form, while the other is in the *syn* configuration. After a second round of DNA replication, one DNA duplex will have a transversion at that point (*blue*).

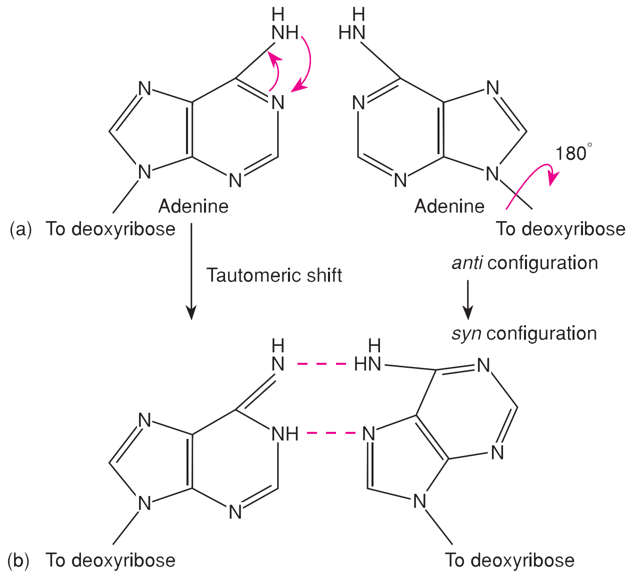


**Figure 12.18** Transversion mutagenesis. An AA base pair can form if one base undergoes a tautomeric shift while the other rotates about its glycosidic (sugar) bond. In (*a*), both bases are in their normal configurations; no hydrogen bonding occurs. In (*b*), hydrogen bonds are possible.

Some base-pair mutations can have serious results. If guanine undergoes an oxidation to 8-oxoguanine (fig. 12.19), it pairs with adenine. A GC base pair is converted to a TA base pair through an 8-oxoguanine-adenine intermediate. This transversion has been found to be common in cancers.

Since 1953, when Watson and Crick first described the structure of DNA, tautomerization has been accepted as the obvious source of most transition mutations. However, recent structural data has cast some doubt on this assumption. X-ray crystallography and nuclear magnetic resonance (NMR) studies indicate that both bases in transition mismatches may be in their normal forms. Other mechanisms, similar to wobble base pairing (see chapter 11), may be responsible for most transition mutations. These studies also indicate that some transversions result from direct purine-purine or pyrimidine-pyrimidine base pairing during DNA synthesis. Much work needs to be done to clarify the nature of spontaneous mutagenesis.

## Chemical Mutagenesis

Muller demonstrated that X rays can cause mutation. Certain chemical and temperature treatments can also cause mutation. Determining the mode of action of various chemical mutagens has provided insight into the mutational process as well as the process of carcinogenesis (box 12.1). In addition, knowing how chemical mutagens act has allowed geneticists to produce large numbers of specific mutations at will (box 12.2).

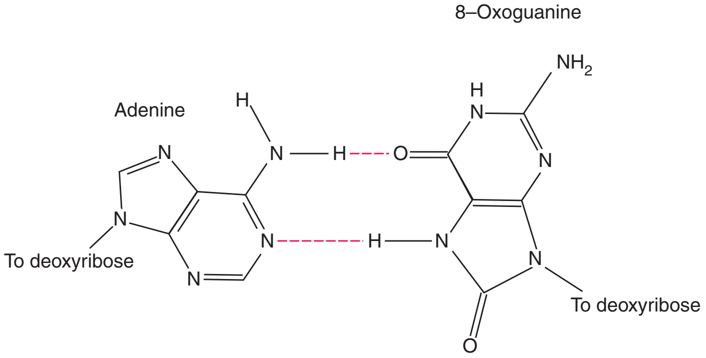


**Figure 12.19** The structure of 8-oxoguanine, which base pairs with adenine, converting a GC base pair to a TA base pair—a transversion. (After C. Mol, et al., "DNA repair mechanisms for the recognition and removal of damaged DNA bases," *Annual Review of Biophysics and Biomolecular Structure,* 28:101–28,1999, figure 6.)

### *Transitions*

Transitions are routinely produced by base analogues. Two of the most widely used base analogues are the pyrimidine analogue 5-bromouracil (5BU) and the purine analogue 2-aminopurine (2AP; fig. 12.20). The mutagenic mechanisms of the two are similar. The 5-bromouracil is incorporated into DNA in place of thymine; it acts just like thymine in DNA replication and, since it doesn't alter the hydrogen bonding, should induce no mutation. However, it seems that the bromine atom causes 5-bromouracil to tautomerize more readily than thymine does. Thus, 5-bromouracil goes from the keto form (fig. 12.20) to the enol form more readily than thymine. Transitions frequently result when the enol form of 5-bromouracil pairs with guanine.

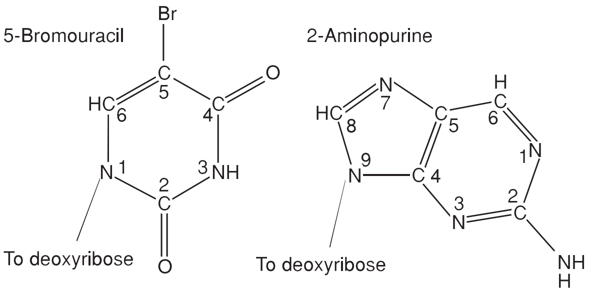

5-Bromouracil
To deoxyribose
2-Aminopurine
To deoxyribose

**Figure 12.20** Structure of the base analogues 5-bromouracil (5BU) and 2-aminopurine (2AP).

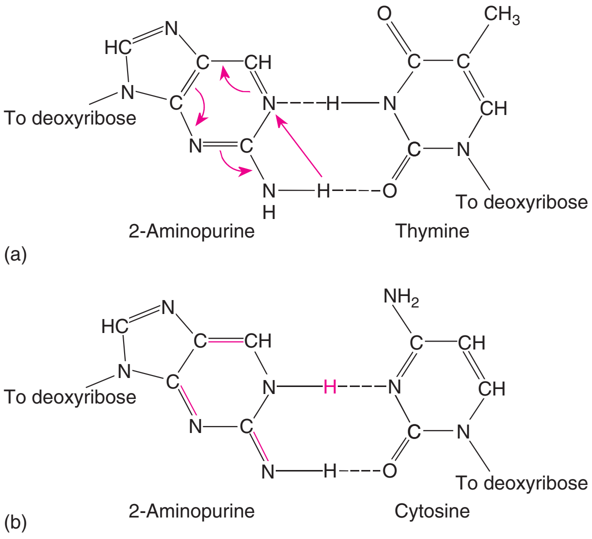

To deoxyribose
2-Aminopurine
Thymine
To deoxyribose
(a)
To deoxyribose
2-Aminopurine
Cytosine
To deoxyribose
(b)

**Figure 12.21** Two possible base pairs with 2-aminopurine. (*a*) In the normal state, 2-aminopurine acts like adenine and pairs with thymine. (*b*) In the rare state, 2-aminopurine acts like guanine and forms complementary base pairs with cytosine.

The 2-aminopurine is mutagenic by virtue of the fact that it can, like adenine, form two hydrogen bonds with thymine. When in the rare state, it can pair with cytosine (fig. 12.21). Thus, at times it replaces adenine, and at other times guanine. It promotes transition mutations.

Nitrous acid ($HNO_2$) also readily produces transitions by replacing amino groups on nucleotides with keto groups ($—NH_2$ to $=O$). The result is that cytosine is converted to uracil, adenine to hypoxanthine, and guanine to xanthine. As figure 12.22 shows, transition

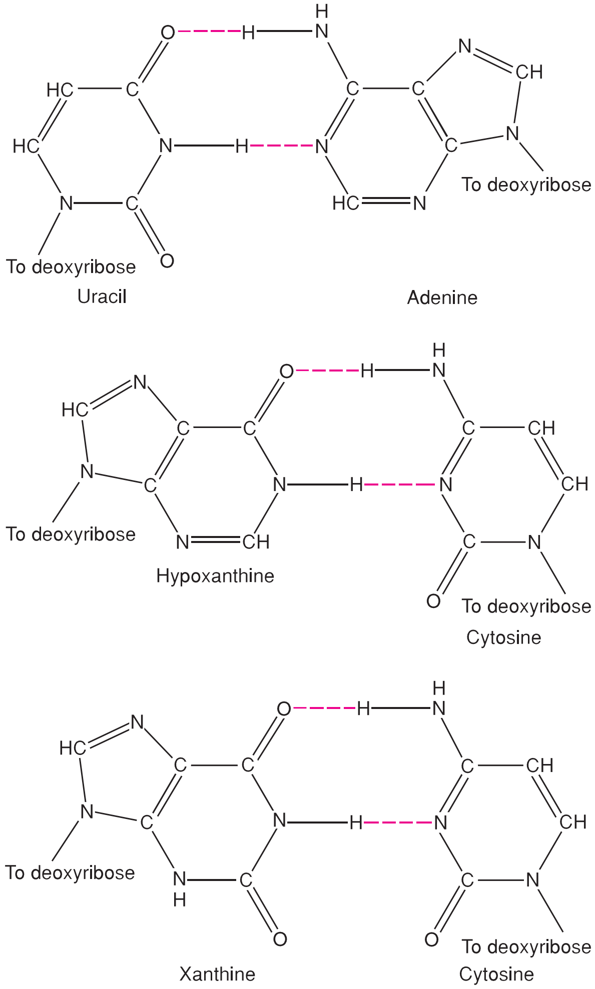

To deoxyribose
Uracil
To deoxyribose
Adenine
To deoxyribose
Hypoxanthine
To deoxyribose
Cytosine
To deoxyribose
Xanthine
To deoxyribose
Cytosine

**Figure 12.22** Nitrous acid converts cytosine to uracil, adenine to hypoxanthine, and guanine to xanthine. Uracil pairs with adenine, whereas its progenitor, cytosine, normally pairs with thymine; hypoxanthine pairs with cytosine, whereas its progenitor, adenine, normally pairs with thymine; and xanthine pairs with cytosine—the same base that guanine, its progenitor, pairs with. Thus, only the first two pairings result in transition mutations.

## BOX 12.1

## Biomedical Applications

### *The Ames Test for Carcinogens*

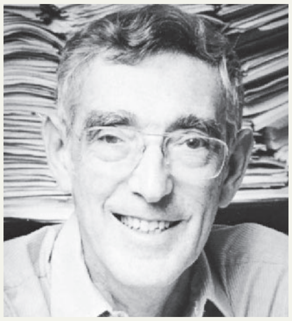
Which chemicals cause cancer in human beings? It is difficult to determine whether any substance is a carcinogen. Tests for carcinogenicity usually involve administering the substance in question to laboratory rats or mice to determine whether the substance actually causes cancer in these animals. Even these tests, however, are not absolute predictors of cancer in people. Tests of this nature are very expensive ($1 million to $2 million each) and time-consuming (three to four years). Since more than fifty thousand different chemical compounds are used in industry, with thousands more added each year, the challenge of making the working environment as well as the general environment safe seems overwhelming. We can, however, make a preliminary determination about the cancer-causing properties of any substance very quickly because of the relationship between mutagenicity and carcinogenicity. Many substances in the environment that can cause cancer also cause mutations. Both kinds of effects are related to DNA damage.

Bruce Ames, at the University of California at Berkeley, developed a routine screening test for mutagenicity. Substances that prove positive in this test are suspected of being carcinogens and would have to be tested further to determine their potential to cause cancer in mammals.

Ames worked with a strain of *Salmonella typhimurium* that requires histidine to grow. This strain will not grow on minimal medium. However, the strain will grow if a mutagen is added to the medium, causing the defective gene in the histidine pathway to revert to the wild-type. (Mutagens inducing gross chromosomal damage, such as deletions or inversions, will not be detected.) Under normal circumstances, there is a background mutation rate; a certain number of *Salmonella* cells revert spontaneously, and therefore a certain number of colonies will grow on the minimal medium. A mutagen, however, increases the number of colonies that can grow on minimal medium. This procedure is, therefore, a rapid, inexpensive, and easy test for mutagenicity.

To improve this test's ability to detect carcinogens, Ames added a supplement of rat liver extract to the medium. It is known that, although many substances are themselves not carcinogens, the breakdown of these substances in the liver creates substances that are carcinogenic. Rat liver enzymes act on a substance the same way human livers do, converting a noncarcinogenic primary substance into a possible carcinogen. The liver enzymes can also make a mutagen nonmutagenic.

Other short-term tests are in use that effectively duplicate the Ames test. These include tests for mutagenicity in mouse lymphoma cells and two tests in Chinese hamster ovary cells: a test for chromosomal aberrations and a test for sister-chromatid exchanges. None of these tests surpasses the Ames test, which has scored better than 90% correct when tested with hundreds of known carcinogens. Thousands of other substances have been subjected to this test; many have proven to be mutagenic. These substances are usually withdrawn from the workplace or home environment. From time to time, we read that a certain substance is believed to be carcinogenic and is being removed from grocery store shelves. Examples have included hair dyes, food preservatives, food-coloring agents, and artificial sweeteners. Many of these first were suspected after they failed the Ames test.



Bruce Ames (1928– ).
(Courtesy Dr. Bruce Ames.)

## BOX 12.2

### Experimental Methods

#### *In Vitro Site-Directed Mutagenesis*

One way of studying the way that proteins work is to change the sequence of amino acids in the protein. For example, if a scientist were working on the active site of a particular enzyme, he or she could learn how the enzyme modifies its substrates by changing one or a few amino acids. Changes could be made, for example, in order to study the role of shape or charge on the functioning of the enzyme. Advances in recombinant DNA techniques have made it possible for a research scientist to create exactly the changes he or she wants in a protein.

To begin with, the gene for the protein or enzyme must be cloned so that it can be manipulated (see chapter 13). Once cloned, deletions are easy to create with restriction endonucleases (described in chapter 13). If a particular endonuclease cuts the gene in two places, the intervening segment can be spliced out (fig. 1*a*; see chapter 13). If the endonuclease cuts only once, exonucleases can digest the ends of the cut, extending the deletion away from the cut in both directions (fig. 1*b*). Insertions can be created by either cutting the gene and repairing the single-stranded ends (fig. 1*c*) or by creating an oligonucleotide (a linker) with the desired sequence and inserting the linker at the site of an endonuclease cut (fig. 1*d*).

Far more impressive, however, is the ability to change a single specific codon in order to replace any amino acid in the protein with any other amino acid. The process involves directed mutagenesis using artificially created oligonucleotides.

Basically, a short sequence of DNA (an oligonucleotide) is synthesized complementary to a region of the cloned gene, but with a change in one or more bases of a codon to specify a different amino acid. That oligonucleotide is then hybridized with the single-stranded form of the clone (fig. 2). Although one or more

*continued*

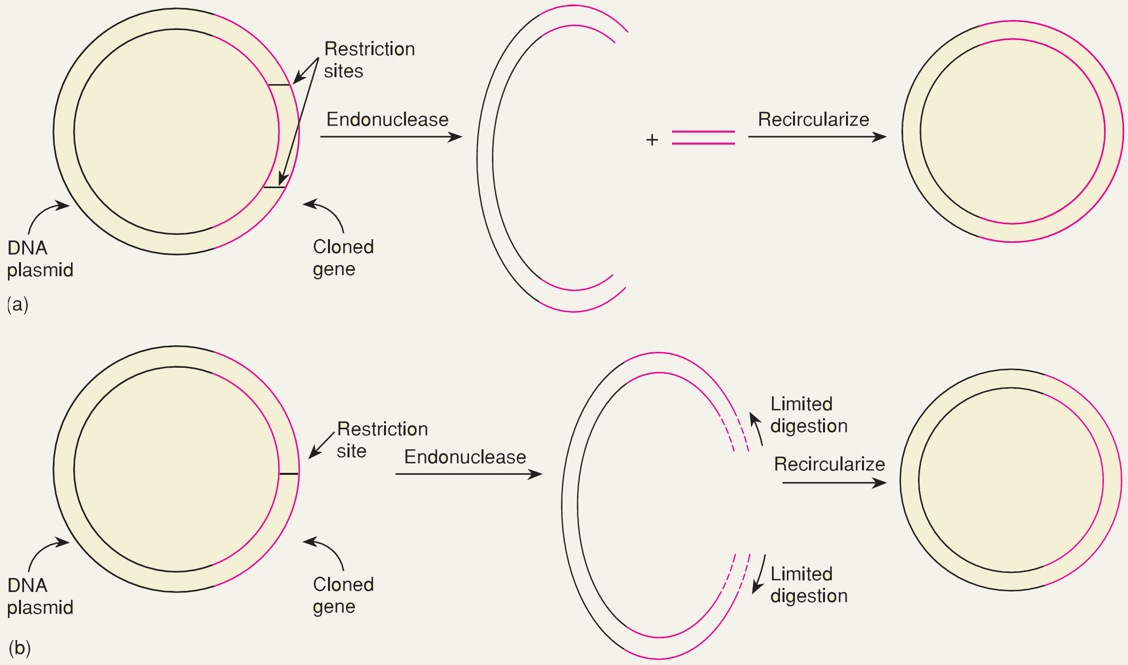


**Figure 1** A cloned gene can be mutated in several ways. (*a*) If a restriction endonuclease has two sites in the gene, the intermediate piece can be spliced out. (*b*) If the endonuclease has only one site, the gene can be opened at that site, and limited digestion by exonucleases will delete part of the gene.

*continued*

## BOX 12.2 CONTINUED

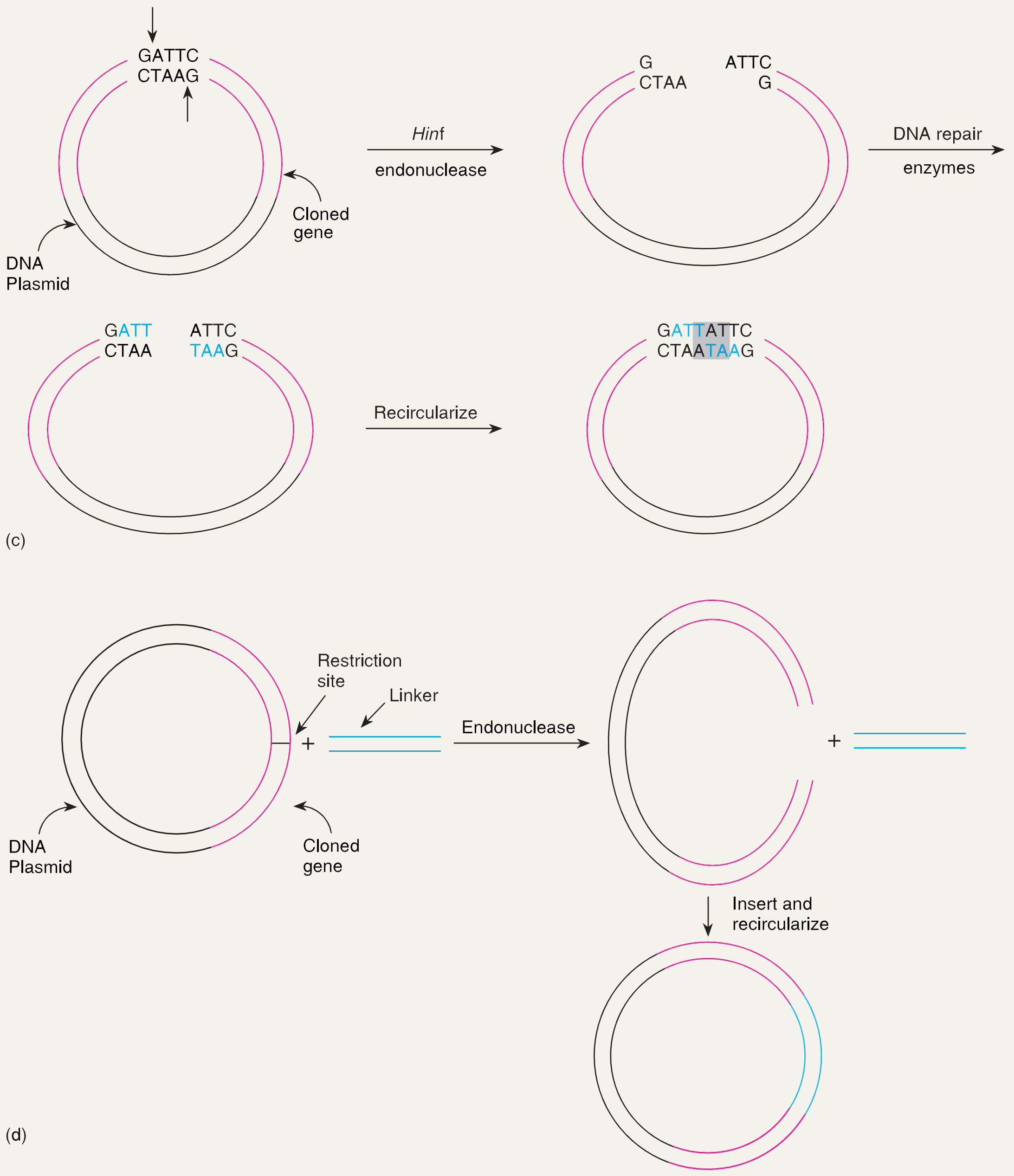

GATTC
CTAAG
Cloned gene
DNA Plasmid
Hinf endonuclease
G
CTAA
ATTC
G
DNA repair enzymes
GATT
CTAA
ATTC
TAAG
Recircularize
GATTATTC
CTAATAAG
(c)
Restriction site
Linker
+
Endonuclease
+
DNA Plasmid
Cloned gene
Insert and recircularize
(d)

**Figure 1—*continued*** (*c*) If an endonuclease has an offset region between its splice points of three or six nucleotides (one or two codons), that length can be inserted by repairing the single-stranded ends after cutting by the endonuclease. The resulting blunt ends can be spliced together. (Note that actually an ATT region has been converted to an ATTATT region. If reading codons along the DNA, the actual insertion is of a TAT codon.) (*d*) A linker of any length (usually the length of a specific number of codons) can be inserted at a restriction site.

bases will not match, hybridization can usually be facilitated by adjusting the pH or ionic strength of the solution. The hybridized oligonucleotide is then used as a primer for DNA replication; the whole plasmid is replicated, resulting in hybrid DNA. In subsequent DNA replications of the hybrid, both the original gene and the mutated DNA will be produced. The latter can be isolated by appropriate selection methods; it is a plasmid with a cloned gene that has the exact mutation the researcher wanted. Using techniques of this type, geneticists have made many advancements in understanding exactly how various components of an enzyme contribute to its function.

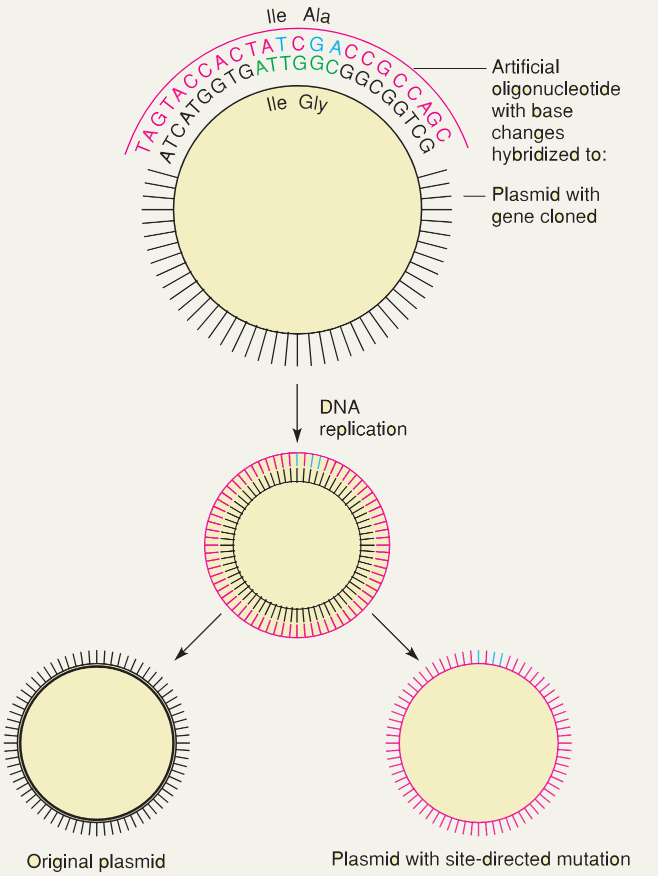

Ile Ala
TAGTACCACTATCGACCGCCAGC
ATCATGGTGATTGGCGGCGGTCG
Ile Gly
Artificial oligonucleotide with base changes hybridized to:
Plasmid with gene cloned
DNA replication
Original plasmid
Plasmid with site-directed mutation

**Figure 2** Site-directed mutagenesis can involve any nucleotide(s). In this case, an inserted gene with an Ile-Gly sequence is converted, at the direction of the investigator, to an Ile-Ala sequence. A single-stranded form of the plasmid is isolated. A synthetically prepared oligonucleotide (twenty-three bases in this example) is added. It can be made to hybridize at the complementary site despite differing by three bases. Then DNA replication is carried out using the oligonucleotide configuration as a primer. After the strands of the duplex are separated, the investigators can isolate the original plasmid as well as the mutated plasmid. (Note that the investigators changed two codons, although they changed only one amino acid, because they also wanted to introduce an Alu site at that point for future studies.) (From J. E. Villafranca, et al., "Directed mutagenesis of dihydrofolate reductase," *Science* 222:782–88. Copyright 1983 by the AAAS.)

mutation results from two of the changes. Uracil pairs with adenine instead of guanine, thus leading to a UA base pair in place of a CG base pair; hypoxanthine (H) pairs with cytosine instead of thymine, the original base paired with adenine. Thus, in this case, an HC base pair replaces an AT base pair. Both of these base pairs (UA and HC) are transition mutations. Xanthine, however, pairs with cytosine just as guanine does. Thus, the replacement of guanine with xanthine does not cause changes in base pairing.

Like nitrous acid, heat can also deaminate cytosine to form uracil and thus bring about transitions (CG to TA). Apparently, heat can also bring about transversions by an unknown mechanism.

### *Transversions*

Ethyl methane sulfonate ($CH_3SO_3CH_2CH_3$) and ethyl ethane sulfonate ($CH_3CH_2SO_3CH_2CH_3$) are agents that cause the removal of purine rings from DNA. The multistep process begins with the ethylation of a purine ring and ends with the hydrolysis of the glycosidic (purine-deoxyribose) bond, causing the loss of the base. These sites where this happens are referred to as AP (apurinic-apyrimidinic) sites. If the AP site is not repaired, any of the four DNA bases could be inserted into the new strand opposite the gap (fig. 12.23). If thymine is placed in the newly formed strand, then the original base pair is restored; insertion of cytosine results in a transition mutation; insertion of either adenine or guanine results in a transversion mutation. Of course, the gap is still there, and it continues to generate new mutations each generation until it is repaired. During DNA replication in *E. coli,* the polymerase tends to place adenine opposite the gap more frequently than it places other bases.

### *Insertions and Deletions*

The molecules of the acridine dyes, such as proflavin and acridine orange (fig. 12.24), are flat. Presumably, they initiate mutation by inserting into the DNA double helix, causing the helix to buckle in the region of insertion, possibly leading to base additions and deletions during DNA replication. Crick and Brenner used acridine-induced mutations to demonstrate both that the genetic

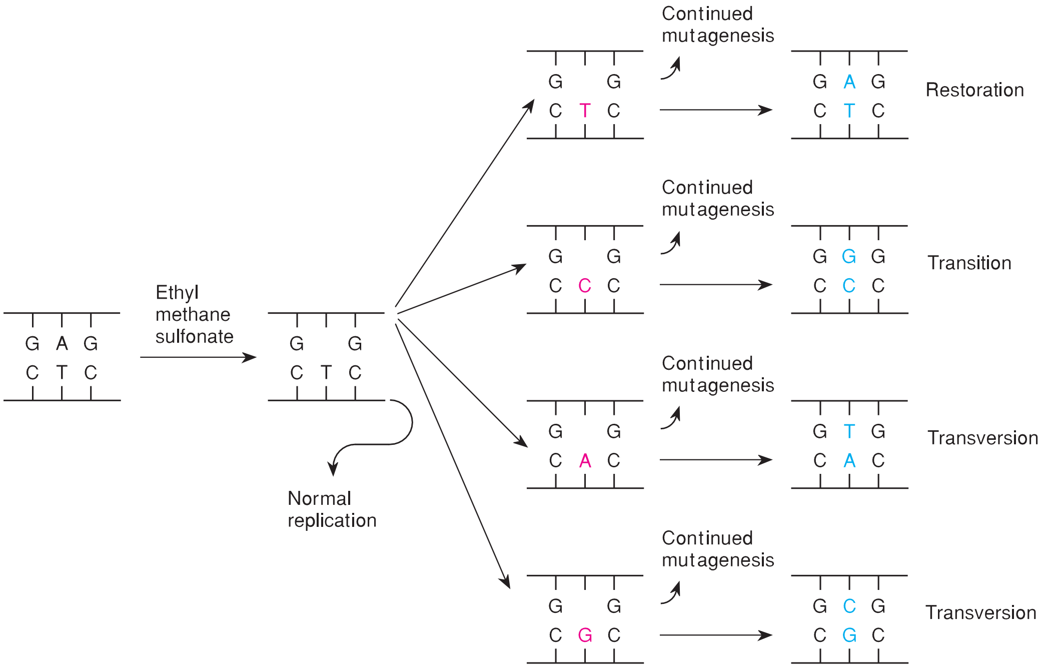


**Figure 12.23** Four possible outcomes after treatment of DNA with an alkylating agent, which removes the purine—adenine in this example. The bases shown in *red* are the four bases that DNA polymerase may insert opposite the gap. After another round of DNA replication, the gap remains to generate further mutations. The inserted base forms a base pair (*blue*), which can be a restoration or a transition or transversion mutation.

code was read from a fixed point and that it was triplet (chapter 11).

## *Misalignment Mutagenesis*

Additions and deletions in DNA can also come about by misalignment of a template strand and the newly formed (progeny) strand in a region containing a repeated sequence. For example, in figure 12.25 we expect the progeny strand to contain six adjacent adenines because the template strand contains six adjacent thymines. Misalignment of the progeny strand results in seven consecutive adenines: six thymines replicated, plus one already replicated but misaligned. Misalignment of the template strand results in five consecutive adenines because one thymine is not available in the template. Regions with long runs of a particular base may be very mutation prone. They may explain the "hot spots" observed by Benzer (see fig. 12.8) and others.

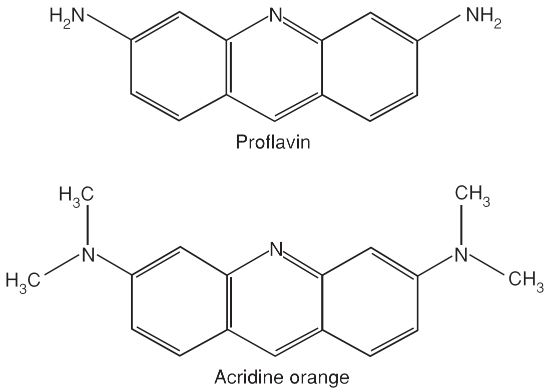

H2N
N
NH2
Proflavin
H3C
CH3
H3C
N
N
N
CH3
Acridine orange

**Figure 12.24** Structure of two acridine dyes: proflavin and acridine orange.

## *Intergenic Suppression*

When a critical mutation occurs in a codon, several routes can still lead to survival of the individual; simple reversion and intragenic suppression are two that we have already considered. A third route is through **intergenic suppression**—restoration of the function of a mutated gene by changes in a different gene, called a **suppressor gene.** Suppressor genes are usually transfer RNA genes. When mutated, intergenic suppressors change the way in which a codon is read.

Suppressor genes can restore proper reading to nonsense, missense, and frameshift mutations. **Nonsense mutations** convert a codon that originally specified an amino acid into one of the three nonsense

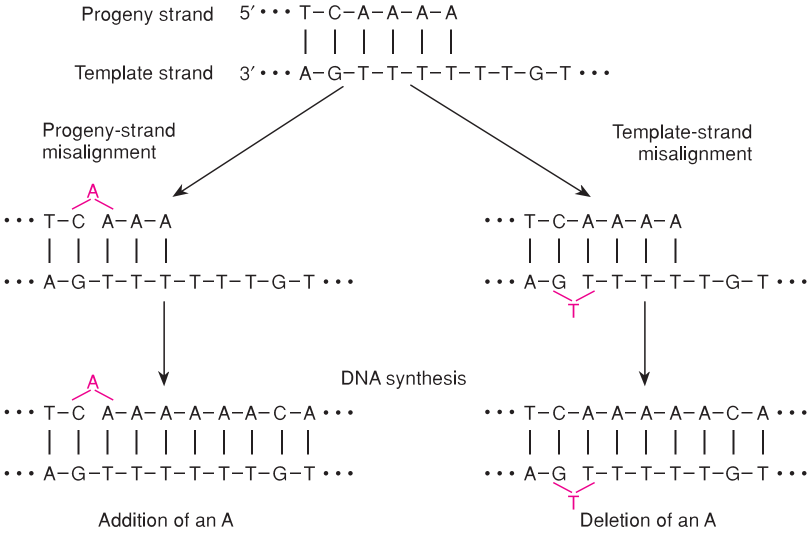
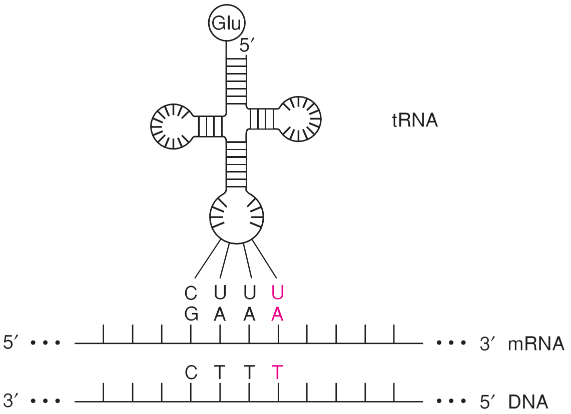

Progeny strand
5′ • • • T–C–A–A–A–A
Template strand
3′ • • • A–G–T–T–T–T–T–T–G–T • • •
Progeny-strand misalignment
Template-strand misalignment
A
• • • T–C A–A–A
• • • A–G–T–T–T–T–T–T–G–T • • •
• • • T–C–A–A–A–A
• • • A–G T–T–T–T–T–G–T • • •
T
DNA synthesis
A
• • • T–C A–A–A–A–A–A–C–A • • •
• • • A–G–T–T–T–T–T–T–G–T • • •
• • • T–C–A–A–A–A–A–C–A • • •
• • • A–G T–T–T–T–T–G–T • • •
T
Addition of an A
Deletion of an A

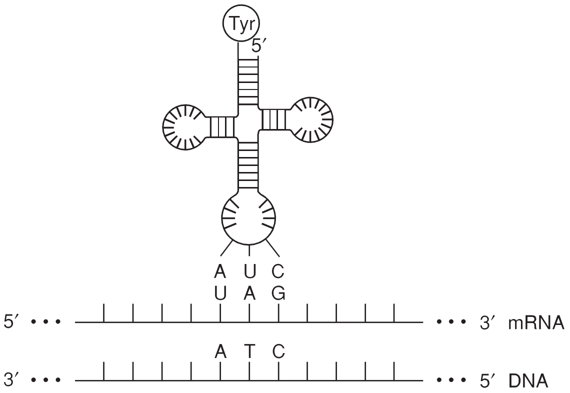
**Figure 12.25** Misalignment of a template or progeny strand during DNA synthesis. If the progeny strand is misaligned after DNA replication has begun, the resulting progeny strand will have an additional base. If the template strand is misaligned during DNA replication, the resulting progeny strand will have a deleted base. These changes will show up after another round of DNA replication. (From J.W. Drake, B.W. Glickman, and L.S. Ripley, "Updating the Theory of Mutation," *American Scientist,* 71:621–630, 1983. Reprinted with permission of American Scientist, magazine of Sigma Xi, The Scientific Research Society.)

codons. **Missense mutations** change a codon so that it specifies a different amino acid. Frameshift mutations, by additions or deletions of nucleotides, cause an alteration in the reading frame of codons. A frameshift mutation, caused by the insertion of a single base, can be suppressed by a transfer RNA that has an added base in its anticodon (fig. 12.26*a*). It reads four bases as a codon and thus restores the original reading frame.

The transfer RNA produced by a nonsense suppressor gene reads the nonsense codon as if it were a codon for an amino acid; an amino acid is placed into the protein, and reading of the messenger RNA continues. At least three suppressors of the mutant amber codon (UAG) are known in *E. coli.* One suppressor puts tyrosine, one puts glutamine, and one puts serine into the protein chain at the point of an amber codon. Normally, tyrosine transfer RNA has the anticodon 3′-AUG-5′. The suppressor transfer RNA that reads amber as a tyrosine codon has the anticodon 3′-AUC-5′, which is complementary to amber. Hence, a mutated tyrosine transfer RNA reads amber as a tyrosine codon (fig. 12.26*b*).



(a) Frameshift suppression



(b) Nonsense suppression

**Figure 12.26** Frameshift and nonsense suppression by mutant transfer RNAs. In (*a*), a thymine has been inserted into DNA, resulting in a frameshift. However, a transfer RNA with four bases in the codon region reads the inserted base as part of the previous codon in the messenger RNA. The frameshift thus does not occur. In (*b*), an *amber* mutation (UAG), which normally results in chain termination, is read as tyrosine by a mutant tyrosine transfer RNA that has the anticodon sequence complementary to the *amber* codon.

If the amber nonsense codon is no longer read as a stop signal, then won't all the genes terminating in the amber codon continue to be translated beyond their ends, causing the cell to die? In the tyrosine case, two genes for tyrosine transfer RNA were found; one contributes the major fraction of the transfer RNAs, and the other, the minor fraction. It is the minor-fraction gene that mutates to act as the suppressor. Thus, most messenger RNAs are translated normally, and most amber mutations result in premature termination, although a sufficient number are translated (suppressed) to ensure the viability of the mutant cell. In general, intergenic suppressor mutants would be eliminated quickly in nature because they are inefficient—the cells are not healthy. In the laboratory, we can provide special conditions that allow them to be grown and studied.

## Mutator and Antimutator Mutations

Whereas intergenic suppressors represent mutations that "restore" the normal phenotype, mostly through mutation of transfer RNA loci, **mutator** and **antimutator mutations** cause an increase or decrease in the overall mutation rate of the cell. They are frequently mutations of DNA polymerase, which, as you remember, not only polymerizes DNA nucleotides 5′ → 3′ complementary to the template strand, but also checks to be sure that the correct base was put in (they proofread). If, in the proofreading process, the polymerase discovers an error, it can correct this error with its 3′ → 5′ exonuclease activity. Mutator and antimutator mutations sometimes involve changes in the polymerase's proofreading ability (exonuclease activity).

Phage T4 has its own DNA polymerase with known mutator and antimutator mutants. Mutator mutants are very poor proofreaders (they have low exonuclease-to-polymerase ratios), and thus they introduce mutations throughout the phage genome. Antimutator mutants, however, have exceptionally efficient proofreading ability (high exonuclease-to-polymerase ratios) and, therefore, result in a very low mutation rate for the whole genome (box 12.3).

## BOX 12.3

### Experimental Methods

### *Adaptive Mutation*

Although we discuss evolution in detail at the end of the book, note here that we view mutation as a random process, not one that occurs because a cell "needs" a particular mutation. For example, Luria and Delbrück's work showed that the mutation in *E. coli* for resistance to phage T1 occurred randomly before exposure to the phage, not because the cells would benefit from the mutation.

Because of the entrenched dogma that mutations occur through random processes, the scientific community was startled when, in 1988, John Cairns—a highly respected senior scientist—and colleagues reported a new observation—**adaptive, or directed, mutation,** occurring when the cell needed it. Their system was the *lacZ* gene in *E. coli.* Cells that could not use lactose as an energy source ($lacZ^-$) were plated on a medium in which lactose was the sole energy source. Some mutants already there of course produced colonies ($lacZ^+$). The expectation was that no new mutations would occur over time because the $lacZ^-$ cells would have either died or stopped metabolizing. Unexpectedly, Cairns and colleagues found that more and more colonies appeared over time, coming from cells that had mutated to $lacZ^+$. As a control, they looked for revertants of other genes not involved with lactose metabolism. These mutations did not occur in a directed manner.

Scientists were extremely skeptical of this work for two reasons. First, it seemed to fly in the face of our common understanding of the mutational process. Second, there were no obvious explanations for how this could occur. Numerous articles were published refuting the notion of adaptive mutations and suggesting other explanations for the results. These explanations included artifacts of miscounting cells to mutants that were extremely slow growing, but there all the time.

In the past several years, other scientists have found at least a half dozen similar results in other organisms and other genes. The debate is ongoing, but work published in mid-1994 seems to have recast it into the realm of methods of mutagenesis rather than non-Darwinian processes. Several scientists found that the "directed" mutations seem to be of a certain type, mainly single nucleotide deletions within runs of the same nucleotide: for example, a deletion of a C in a CCCC sequence. This type of error happens during DNA replication and could be the result of a repair deficiency. That is, under extreme duress, the cells may be going into a "hypermutational" mode, or selection may favor hypermutable genotypes in which repair mechanisms are shut down in order to intentionally create lots of errors in the DNA. Any errors that do not alleviate the problem result in cell death, a death that was inevitable anyway; however, some errors will correct the problem ($lacZ^-$ to $lacZ^+$). Those cells will survive. One recent study indicated that a subpopulation of about 0.06% of the population was hypermutable, with a mutation rate about 200 times that of the normal cells. That group of cells could account for the adaptive mutations. Other scientists found that when the locus of importance was on a plasmid, the adaptive mutation could occur by increased replication of the plasmid. (Currently, the favored term is *adaptive mutation,* rather than *directed mutation,* indicating that the mutations are useful to the cell, not that some unknown process directs them.)

This controversy, although not necessarily resolved, has actually brought out the best in the scientific method: A skeptical scientific community tried its best to refute an unreasonable observation. Other scientists then repeated the observation and extended it, making it more worthy of further study. Finally, further work has given us reasonable mechanisms that not only require no rejection of the original concept of random mutation, but actually give us hypotheses to test further.

## DNA REPAIR

Radiation, chemical mutagens, heat, enzymatic errors, and spontaneous decay constantly damage DNA. For example, it is estimated that several thousand DNA bases are lost each day in every mammalian cell due to spontaneous decay. Some types of DNA damage interfere with DNA replication and transcription. In the long evolutionary challenge to minimize mutation, cells have evolved numerous mechanisms to repair damaged or incorrectly replicated DNA. Many enzymes, acting alone or in concert with other enzymes, repair DNA. Repair systems are generally placed in four broad categories: damage reversal, excision repair, double-strand break repair, and postreplicative repair. Enzymes that process repair steps have been conserved during evolution. That is, enzymes found in *E. coli* have homologues in yeast, fruit flies, and human beings. However, eukaryotic systems are almost always more complex.

## Damage Reversal

Ultraviolet (UV) light causes linkage, or **dimerization,** of adjacent pyrimidines in DNA (fig. 12.27). Although cytosine-cytosine and cytosine-thymine dimers are occasionally produced, the principal products of UV irradiation are thymine-thymine dimers. These can be repaired in several different ways. The simplest is to reverse the dimerization process and restore the original unlinked thymines.

In *E. coli,* an enzyme called DNA photolyase, the product of the *phr* gene (for **photoreactivation**), binds to dimerized thymines. When light shines on the cell, the enzyme breaks the dimer bonds with light energy. The enzyme then falls free of the DNA. This enzyme thus reverses the UV-induced dimerization. Another example of an enzyme that performs direct DNA repair is $O^6$-mGua DNA methyltransferase, which removes the methyl groups from $O^6$-methylguanine, the major product of DNA-methylating agents (fig. 12.28). Although other repair mechanisms seem to be present in all organisms, photoreactivation is not; it is apparently absent in human beings.

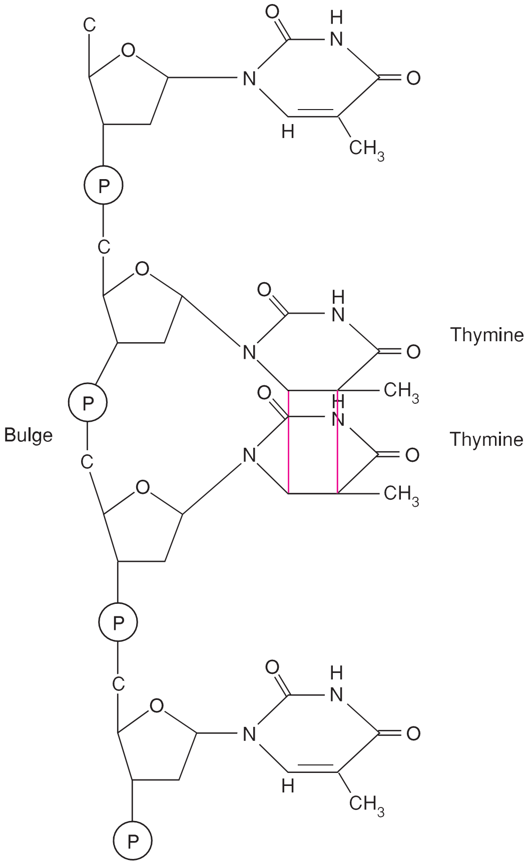


**Figure 12.27** UV-induced dimerization of adjacent thymines in DNA. The *red lines* represent the dimer bonds in the adjacent thymines.

## Excision Repair

**Excision repair** refers to the general mechanism of DNA repair that works by removing the damaged portion of a DNA molecule. Various enzymes can sense damage or distortion in the DNA double helix. During excision repair, bases and nucleotides are removed from the damaged strand. The gap is then patched using complementarity with the remaining strand. We can broadly categorize these systems as base excision repair and nucleotide excision repair, which includes mismatch repair. We will discuss only the major repair pathways; others exist. Presumably, redundancy in repair has been selected for because of the critical need to keep DNA intact and relatively mutation free.

### Base Excision Repair

A base can be removed from a nucleotide within DNA in several ways: by direct action of an agent such as radiation, by spontaneous hydrolysis, by an attack of oxygen free radicals, or by **DNA glycosylases,** enzymes that sense damaged bases and remove them. Currently, at least five DNA glycosylases are known. For example, uracil-DNA glycosylase, the product of the *ung* gene in

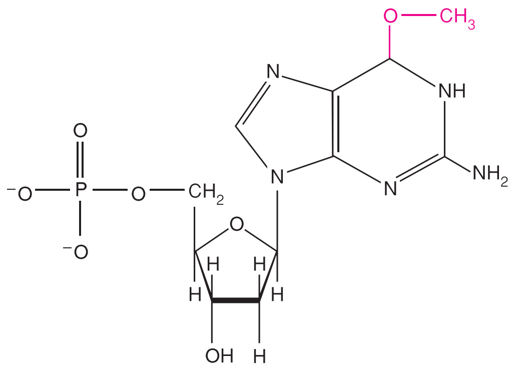


**Figure 12.28** The structure of $O^6$-methylguanine. The *red color* shows the modification of guanine, in which the normal configuration is a double-bonded oxygen (keto form).

*E. coli,* recognizes uracil within DNA and cleaves it out at the base-sugar (glycosidic) bond. The resulting site is called an AP (apurinic-apyrimidinic) site, because of the lack of a purine or pyrimidine at the site (see fig. 12.23). An **AP endonuclease** then senses the minor distortion of the DNA double helix and initiates excision of the single AP nucleotide in a process known as **base excision repair.** The AP endonuclease nicks the DNA at the 5′ side of the base-free AP site. A DNA polymerase then inserts a nucleotide at the AP site; an exonuclease, lyase, or phosphodiesterase enzyme then removes the base-free nucleotide. (Lyases are enzymes that can break C-C, C-O, and C-N bonds.) DNA ligase then closes the nick (fig. 12.29). The replacement of just one base occurs 80–90% of the time. In the remaining 10–20% of cases, several nucleotides may be removed, depending probably on which DNA polymerase (I or III) first repairs the site (fig. 12.29). In mammals, DNA polymerase β performs two roles in base excision repair: It both inserts a new base where the AP site was and also eliminates the AP nucleotide residue by exonuclease activity.

One question that concerned scientists was how the glycosylases gain access to the inappropriate or damaged bases within the double helix. Recently, it has been demonstrated that these enzymes remove the inappropriate or damaged bases by first flipping them out of the interior of the double helix in a process called **base flipping.** For example, the enzyme in human beings that recognizes 8-oxoguanine in DNA (see fig. 12.19), 8-oxoguanine DNA glycosylase, flips the base out to excise it. Base flipping seems to be a common mechanism in repair enzymes that need access to bases within the double helix (fig. 12.30).

### Nucleotide Excision Repair

Whereas base excision repair is initiated by glycosylases and usually involves the replacement of only one nucleotide residue, **nucleotide excision repair** is initiated by enzymes that sense distortions in the DNA backbone and replace a short stretch of nucleotides. For example, six enzymes in *E. coli* excise a short stretch of DNA containing thymine dimers if the dimerization is not reversed by photoreactivation. Two copies of the protein product of the *uvrA* gene (for ultraviolet light—UV—repair) combine with one copy of the product of the *uvrB* gene to form a $UvrA_2UvrB$ complex that moves along the DNA, looking for damage (fig. 12.31). (The complex has 5′ to 3′ helicase activity.) When the complex finds damage such as a thymine dimer, with moderate to large distortion of the DNA double helix, the $UvrA_2$ dimer dissociates, leaving the UvrB subunit alone. This causes the DNA to bend and attracts the protein product of the *uvrC* gene, UvrC. The UvrB subunit first

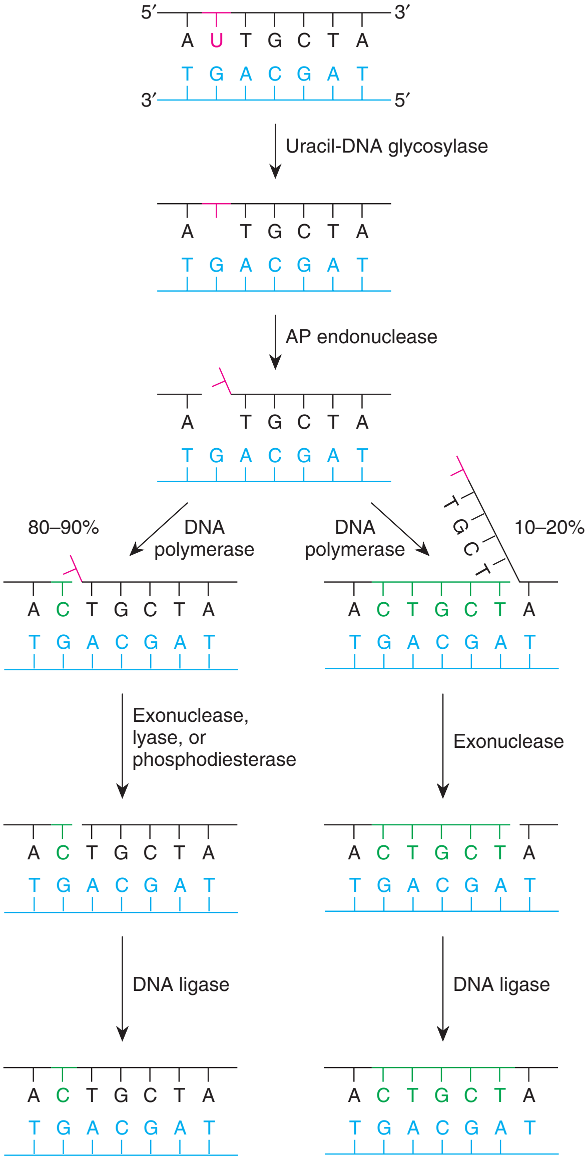
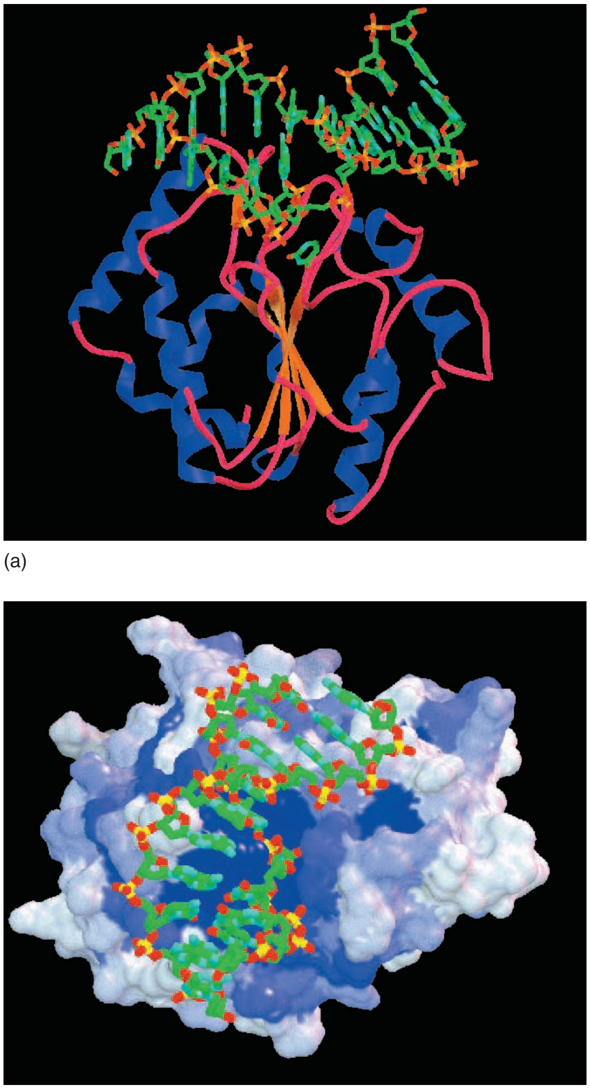


**Figure 12.29** Mechanism of base excision repair. In this case, a uracil-DNA glycosylase enzyme removes a uracil (*red*) from DNA. An AP (apurinic-apyrimidinic) endonuclease nicks the DNA on the 5′ side of the base-free site. Between 80 and 90% of the time, a DNA polymerase will replace the single nucleotide (*green*); an exonuclease, lyase, or phosphodiesterase will remove the base-free nucleotide. The final nick is sealed with DNA ligase. Between 10 and 20% of the time, the DNA polymerase will extend polymerization beyond the single nucleotide. In those cases, an exonuclease and DNA ligase finish the repair. (From T. Lindahl, "The Croonian Lecture, 1996: Endogenous Damage to DNA," *Philosophical Transactions of the Royal Society of London,* B351, pp. 1529–1538, figure 6, 1996.)



(a)

(b)

**Figure 12.30** Two views of the human enzyme, uracil-DNA glycosylase, bound to DNA that has a uracil present. In (*a*), the uracil-containing nucleotide residue has been flipped out and the uracil cleaved; in (*b*), the uracil-containing residue has been flipped out, but the uracil has not yet been cleaved. In both, DNA is a *green* stick figure with *red* oxygen, *yellow* phosphorus, and *blue* nitrogen atoms. In (*a*), the enzyme is shown as a ribbon diagram; note the cleaved uracil bound to the ribbon. In (*b*), the enzyme is shown as a molecular surface; the uracil-containing nucleotide is flipped out by the *purple* knob just left of and below center of the structure with the flipped-out residue to the right. (From S. Parikh, C. Mol, and J. Tainer, "Base excision repair enzyme family portrait" in *Structure*, 1997, 5:1543–1550, fig 1 a&b, p. 1544. Courtesy of J.A. Tainer, The Scripps Research Institute.)

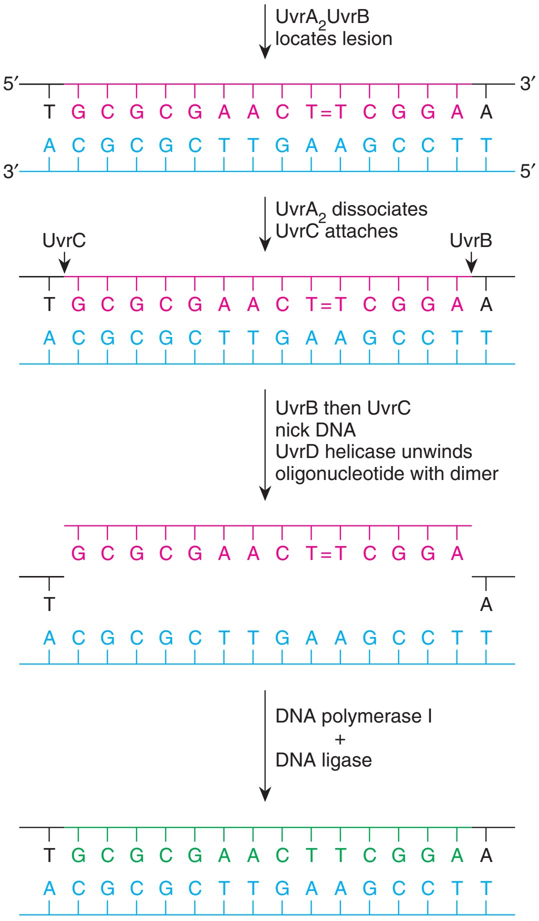

UvrA2UvrB locates lesion
5′
3′
T G C G C G A A C T=T C G G A A
A C G C G C T T G A A G C C T T
3′
5′
UvrA2 dissociates UvrC attaches
UvrC
UvrB
T G C G C G A A C T=T C G G A A
A C G C G C T T G A A G C C T T
UvrB then UvrC nick DNA
UvrD helicase unwinds oligonucleotide with dimer
G C G C G A A C T=T C G G A
T
A
A C G C G C T T G A A G C C T T
DNA polymerase I + DNA ligase
T G C G C G A A C T T C G G A A
A C G C G C T T G A A G C C T T

**Figure 12.31** Nucleotide excision repair. A lesion in DNA (a thymine dimer) is located by a protein made of two copies of UvrA and one of UvrB. Then, the UvrA subunits detach, and UvrC attaches on the 5′ side of the lesion. UvrB nicks the DNA on the 3′ side and UvrC on the 5′ side of the lesion; UvrD helicase unwinds the oligonucleotide containing the lesion (*red*). DNA polymerase I and DNA ligase then repair the patch (*green*).

nicks (hydrolyzes) the DNA four to five nucleotides on the 3′ side of the lesion; next, the UvrC subunit nicks the DNA eight nucleotides on the 5′ side of the lesion. (The three components, UvrA, UvrB, and UvrC, are together called the *ABC excinuclease,* for excision endonuclease.) The enzyme helicase II, the product of the *uvrD* gene, then removes the twelve- to thirteen-base oligonucleotide as well as UvrC. DNA polymerase I fills in the gap and, in the process, evicts the UvrB, and DNA ligase closes the remaining nick (fig. 12.31). This is another relatively simple system designed to detect helix distortions and repair them.

Like base excision repair, nucleotide excision repair is present in all organisms. In yeast, approximately twelve genes are involved, many in what is called the RAD3 group. In human beings, twenty-five proteins are involved; they remove twenty-seven to twenty-nine nucleotides, as compared to twelve to thirteen in *E. coli.*

Transcription and nucleotide excision repair are linked in eukaryotes. Transcription factor TFIIH (see chapter 11) is involved in repair of UV damage; it has helicase activity and is found in both processes. Since it has been shown that genes that are actively being transcribed are preferentially repaired, we can now envision a model in which transcription, when blocked by a DNA lesion like a thymine dimer, signals the formation of a repair complex, using TFIIH in both processes. In prokaryotes, RNA polymerase dissociates from the DNA in this circumstance, losing the nascent transcript. This would be inefficient in eukaryotes, whose genes are much longer and more expensive to transcribe; for example, the human dystrophin gene, defective in the disease Duchenne muscular dystrophy, is 2.4 million bases long and takes almost eight hours to transcribe. We believe that eukaryotic RNA polymerase II backs up when stalled at a DNA lesion and continues after the lesion is repaired, without losing the transcript. Much active research is going on in this area.

In human beings, the autosomal recessive trait *xeroderma pigmentosum* is caused by an inability to repair thymine dimerization induced by UV light. Persons with this trait freckle heavily when exposed to the UV rays of the sun, and they have a high incidence of skin cancer. There are seven complementation groups (loci *XPA-XPG*) whose protein products are involved in the first steps of nucleotide excision repair and whose defects cause xeroderma pigmentosum in human beings. One of them, *XPD,* is a component of TFIIH.

Excision repair triggered by mismatches is referred to as **mismatch repair,** which encompasses about 99% of all DNA repairs. As DNA polymerase replicates DNA, some errors are made that the proofreading polymerase does not correct. For example, a template G can be paired with a T rather than a C in the progeny strand. The GT base pair does not fit correctly in the DNA duplex. The mismatch repair system, which follows behind the replicating fork, recognizes this problem. This system, whose members in *E. coli* are specified by the *mutH, mutL, mutS,* and *mutU* genes, is responsible for the removal of the incorrect base by an excision repair process. (The genes are called *mut* for mutator because mutations of these genes cause high levels of spontaneous mutation in the cells. The *mutU* gene is also known as *uvrD.*) The mismatch repair enzymes initiate the removal of the incorrect base by nicking the DNA strand on one side of the mismatch.

You might wonder how the mismatch repair system recognizes the progeny, rather than the template, base as the wrong one. After all, in a mismatch, there are no defective bases—theoretically, either partner could be the "wrong" base. In *E. coli,* the answer lies in the methylation state of the DNA. DNA methylase, the product of the *dam* locus, methylates 5′-GATC-3′ sequences, which are relatively common in the DNA of *E. coli,* at the adenine residue. Since the mismatch repair enzymes follow the replication fork of the DNA, they usually reach the site of mismatch before the methylase does. Template strands will be methylated, whereas progeny strands, being newly synthesized, will not be. Thus, the methylation state of the DNA cues the mismatch repair enzymes to eliminate the progeny-strand base for repair. After the methylase passes by, both strands of the DNA are methylated, and the methylation cue is gone.

In figure 12.32, we present one model of mismatch repair. The MutS protein, in the form of a homodimer—two

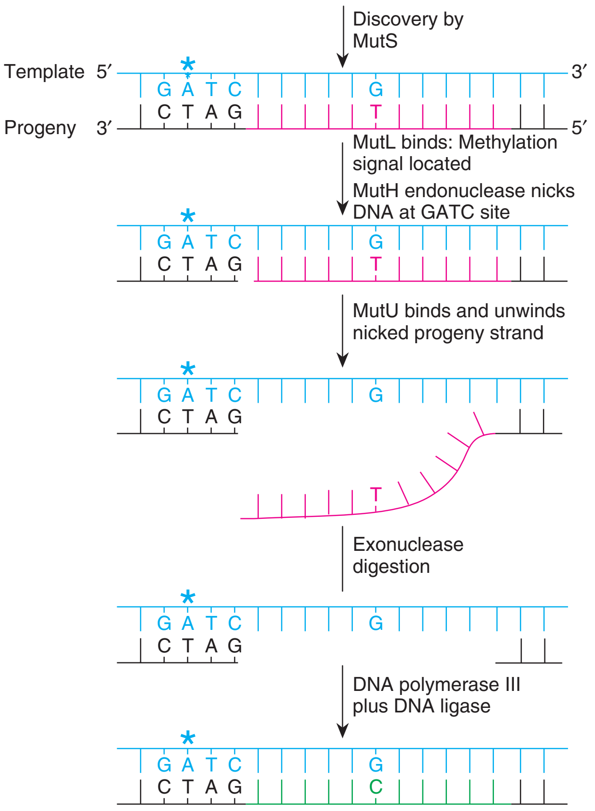

Discovery by MutS
Template 5′
3′
Progeny 3′
5′
*
G A T C
C T A G
G
T
MutL binds: Methylation signal located
MutH endonuclease nicks DNA at GATC site
MutU binds and unwinds nicked progeny strand
Exonuclease digestion
DNA polymerase III plus DNA ligase
C

**Figure 12.32** Mismatch repair. The MutS protein discovers mismatches; MutL binds and the MutH endonuclease nicks the progeny strand at the 3′-CTAG-5′ sequence. MutU helicase unwinds the nicked oligonucleotide with the mismatch (*red*). Exonuclease digestion, followed by DNA polymerase III and DNA ligase repair, completes the operation.

copies of the same protein—finds the mismatch. MutL, also in the form of a homodimer, then binds, and together they find the methylation signal. They also activate the endonuclease MutH, which then nicks the unmethylated strand at the 3′-CTAG-5′ recognition site, which can be one thousand to two thousand bases away from the mismatch. At the recognition site, the MutS-MutL tetramer loads the helicase MutU (UvrD), which then unwinds the nicked strand. Any one of at least four different exonucleases then attacks the unwound oligonucleotide. DNA polymerase III then repairs the gap, and DNA ligase seals it. This sequence of events highlights a common theme in DNA repair: Once a lesion is found, the damaged DNA has some protein bound to it until the repair is finished.

Our understanding of DNA damage and repair helps provide an answer to an evolutionary question—Why does DNA have thymine while RNA has uracil? If we live in an RNA world, in which RNA evolved first, why don't DNA and RNA both contain uracil? One answer is that a common damage to cytosine, spontaneous deamination, results in uracil. If uracil were a normal base in DNA, the conversion of cytosine to uracil by deamination would not leave any clue to a mismatch repair system that a mutation had occurred. Thus, thymine replaces uracil in DNA, since thymine is not confused with any other normal base in DNA by common spontaneous changes. In fact, cytosine, guanine, adenine, and thymine are not converted simply to any other of the bases in DNA. Hence, changes of these bases leave clues for the repair systems.

## Double-Strand Break Repair

Some damage to DNA, such as that caused by ionizing radiation, is capable of breaking both strands of the double helix. When that happens, the cell uses one of two mechanisms to repair the broken ends: It can simply bring the ends back together (a process called *nonhomologous end joining*), or it can use a mechanism that relies on the nucleotide sequences of a homologous piece of DNA, such as a sister chromatid or a homologous chromosome. That method is called *homology-directed recombination.*

In nonhomologous end joining, a protein called Ku, a heterodimer of Ku70 and Ku80, binds to broken chromosomal ends. It then recruits a protein kinase ($PK_{CS}$); their interaction and the interaction with other proteins is stabilized by a scaffold protein called XRCC4 (for X-ray cross complementation group 4). The complex directs the annealing of the broken ends by DNA ligase IV. No particular sequence information is used, and if more than two broken ends are present, incorrect attachments can take place (e.g., translocations). The second method, homology-directed recombination, involves a second piece of DNA homologous to the broken piece. The method is very similar to our current model of DNA recombination and is discussed in the section entitled "Recombination" later in the chapter.

## Postreplicative Repair

When DNA polymerase III encounters certain damage in *E. coli,* such as thymine dimers, it cannot proceed. Instead, the polymerase stops DNA synthesis and, leaving a gap, skips down the DNA to resume replication as far as eight hundred or more bases away. If allowed to remain, this gap will result in deficient and broken DNA. Since part of one strand is absent and the other has damage, there appears to be no viable template for replicating new DNA. However, the cell has two mechanisms to repair this gap: one uses polymerases that can replicate these lesions, and the other is a repair process that uses homologous DNA.

Originally, several proteins were known to facilitate the replication of DNA with lesions; they were believed to interact with the polymerase to make it capable of using damaged DNA as a template. We now know that these proteins are, in fact, polymerases that have the ability to replicate damaged DNA. In *E. coli,* polymerase V can copy damaged DNA. In yeast, polymerases $\eta$ and $\zeta$, also called REV3/7 and RAD30 polymerases, respectively, can also copy damaged DNA. Some of these polymerases are relatively error free; polymerase V and polymerases $\eta$ put adenine-containing nucleotides opposite dimerized thymines. However, polymerases $\zeta$ and the *E. coli* polymerase IV, which also appears during times of damage, are error prone in their replicative roles. One possible reason for this is that the error-prone polymerases developed by evolutionary processes: They create mutations at a time when the cell might need variability. That is, DNA damage can occur when the environment is stressful for the cell; variability might help the cell survive. As we will see later, the cell can sense DNA damage and act appropriately.

In addition to using repair polymerases, the cell can use a second repair mechanism to replicate damaged DNA when the polymerase leaves a gap. A replication fork creates two DNA duplexes. Thus, an undamaged copy of the region with the lesion exists on the other daughter duplex. A group of enzymes, with one specified by the *recA* locus having central importance, repairs the gap. Since the repair takes place at a gap created by the failure of DNA replication, the process is called **postreplicative repair.** The *recA* locus was originally discovered and named in the recombination process. In fact, postreplicative repair is sometimes called recombinational repair, and it shares many enzymes with recombination.

### *The RecA Protein*

The RecA protein has two major properties. First, it coats single-stranded DNA (fig. 12.33) and causes that coated, single-stranded DNA to invade double-stranded DNA (fig. 12.34). By invasion, we mean that the single-stranded DNA attempts to form complementary base pairs with the antiparallel strand of the double-stranded DNA while displacing the other strand of that double helix. A mechanism for this activity, assuming two sites on the enzyme, appears in figure 12.35. RecA continues to move the single-stranded DNA along the double-stranded DNA until a region of homology is found. The second major property of the RecA protein is that, when stimulated by the presence of single-stranded DNA, it causes autocatalysis of another repressor, called LexA, and thus initiates several sequences of reactions.

The RecA protein is responsible for filling a postreplicative gap in newly replicated DNA with a strand from the undamaged sister duplex. Gap-filling

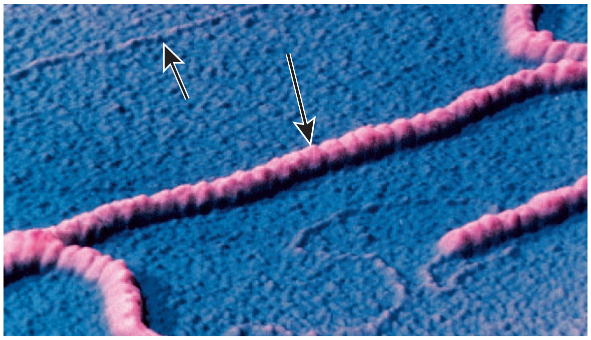


**Figure 12.33** Scanning tunneling microscope picture of single-stranded DNA coated with RecA protein (*large arrow*). The *small arrow* indicates uncoated, double-stranded DNA. (In fig. 12.35, we show how the very large, coated DNA can invade the very small, uncoated DNA.) (© Science VU-IBMRL/Visuals Unlimited.)

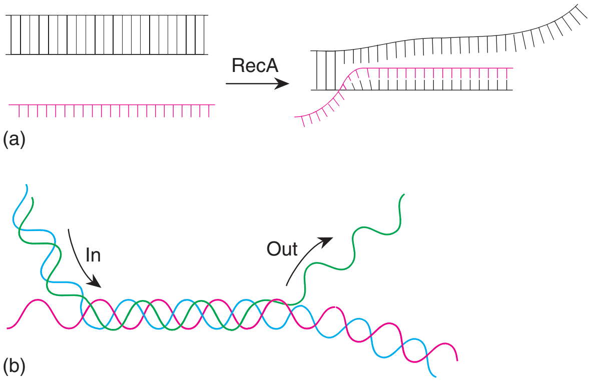


**Figure 12.34** One property of the RecA protein causes single-stranded DNA to invade double-stranded DNA and to move along it until a region of complementarity is found. (*a*) Diagrammatic representation of the invasion of the RecA-coated single-stranded DNA. (*b*) A more realistic diagram of the same event. (Reproduced, with permission, from the *Annual Review of Biochemistry,* Volume 61, © 1992 by Annual Reviews, Inc.)

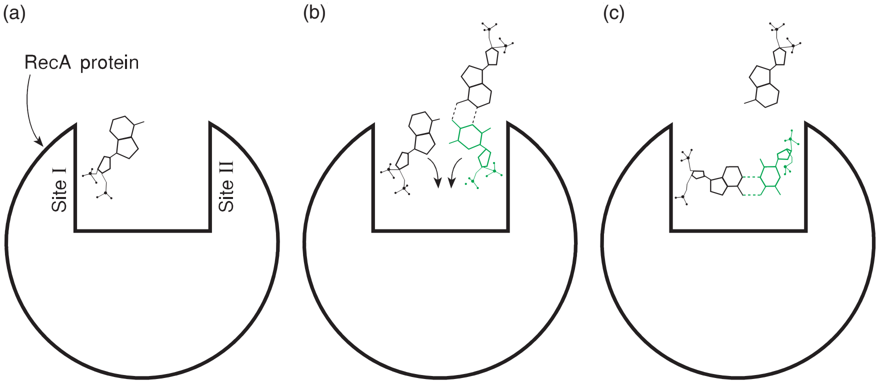


**Figure 12.35** A model of how the RecA protein can cause single-stranded DNA to invade a double-stranded molecule. (*a*) Axial view of one nucleotide (with two phosphate groups) of single-stranded DNA attached at site I in this cross-sectional diagram of the RecA protein. The protein is about 60% larger than actually shown. (*b*) Duplex DNA is bound at site II of RecA. (*c*) RecA protein rotates the bases so that the single-stranded DNA forms a complementary base pair with one strand of the duplex, leaving the other strand of the duplex unpaired (see fig. 12.34). (Reprinted with permission from P. Howard-Flanders, et al., "Role of RecA protein spiral filaments in genetic recombination," *Nature,* 309:215–20. Copyright © 1984 Macmillan Magazines, Limited.)

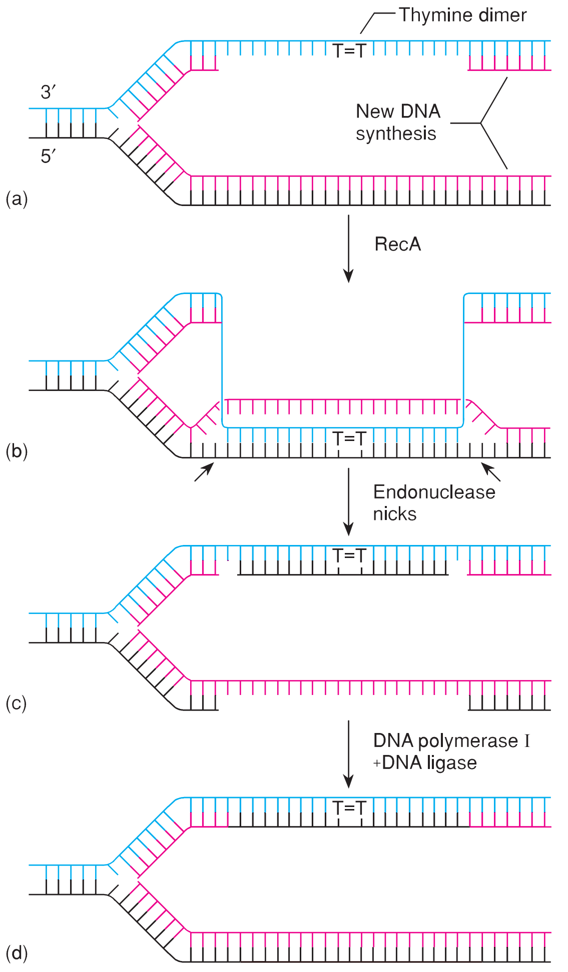


**Figure 12.36** RecA-dependent postreplicative DNA repair. DNA polymerase III skips past a thymine dimer during DNA replication (*a*). With the help of RecA, the single strand with the thymine dimer invades the normal sister duplex (*b*). An endonuclease nicks the new duplex at either side of the thymine dimer site, freeing the new duplex with the thymine dimer and leaving the sister duplex single-stranded (*c*). Repair enzymes then create two intact daughter duplexes (*d*).

processes then complete both strands. In figure 12.36*a,* we see a replication fork with a gap in the progeny strand in the region of a thymine dimer. The RecA protein is responsible for the damaged single strand invading the sister duplex (fig. 12.36*b*). Endonuclease activity then frees the double helix containing the thymine dimer (fig. 12.36*c*). DNA polymerase I and DNA ligase return both daughter helices to the intact state (fig. 12.36*d*). The thymine dimer still exists, but now its duplex is intact, and another cell cycle is available for photoreactivation or excision repair to remove the dimer.

### *The SOS Response*

Postreplicative repair is part of a cell reaction called the **SOS response.** When an *E. coli* cell is exposed to excessive quantities of UV light, other mutagens, or agents that damage DNA (such as alkylating or cross-linking agents), or when DNA replication is inhibited, gaps are created in the DNA. In the presence of this single-stranded DNA, the RecA protein interacts with the LexA protein, the product of the *lexA* gene. The LexA protein normally represses about eighteen genes, including itself. The other genes include *recA, uvrA, uvrB, and uvrD;* two genes that inhibit cell division, *sulA* and *sulB;* and several others. Each of these genes has a consensus sequence in its promoter called the **SOS box:** 5′-CTGX$_{10}$CAG-3′ (where X$_{10}$ refers to any ten bases). The LexA protein normally binds at the SOS box, limiting the transcription of these genes. When single-stranded DNA activates RecA, RecA interacts with the LexA protein to trigger the autocatalytic properties of LexA (fig. 12.37). Transcription then follows from all the genes having an SOS box. The two inhibitors of cell division, the products of the *sulA* and *sulB* genes, presumably increase the amount of time the cell has to repair the damage before the next round of DNA replication.

Eventually, the DNA damage is repaired. There is no single-stranded DNA to activate RecA, and, therefore, LexA is no longer destroyed. LexA again represses the suite of proteins involved in the SOS response, and the SOS response is over. Table 12.5 summarizes some of the enzymes and proteins involved in DNA repair.

As we will mention in chapter 14, λ prophage can be induced into vegetative growth by UV light. This is another effect of the SOS response. RecA not only causes the LexA protein to be inactivated, but also directly inactivates the λ repressor, the product of the λ *cI* gene. From an evolutionary point of view, it makes sense for phage λ to have evolved a repressor protein that the RecA protein inactivates. As a prophage, λ is dependent on the survival of the host cell. When that survival might be in jeopardy, the prophage would be at an advantage if it could sense the danger and make copies of itself that could leave the host. One of these times might be when the host has suffered a lot of DNA damage. The SOS response is a signal to a prophage that the cell has received that damage. Hence, the prophage is induced when RecA acts as a protease; the λ repressor is destroyed, the cro protein becomes dominant, and vegetative growth follows. From an evolutionary perspective, the *E. coli* cell has not created an enzyme (RecA) that seeks out the λ repressor for the benefit of λ. Rather, the λ repressor has evolved for its own advantage to be sensitive to RecA.

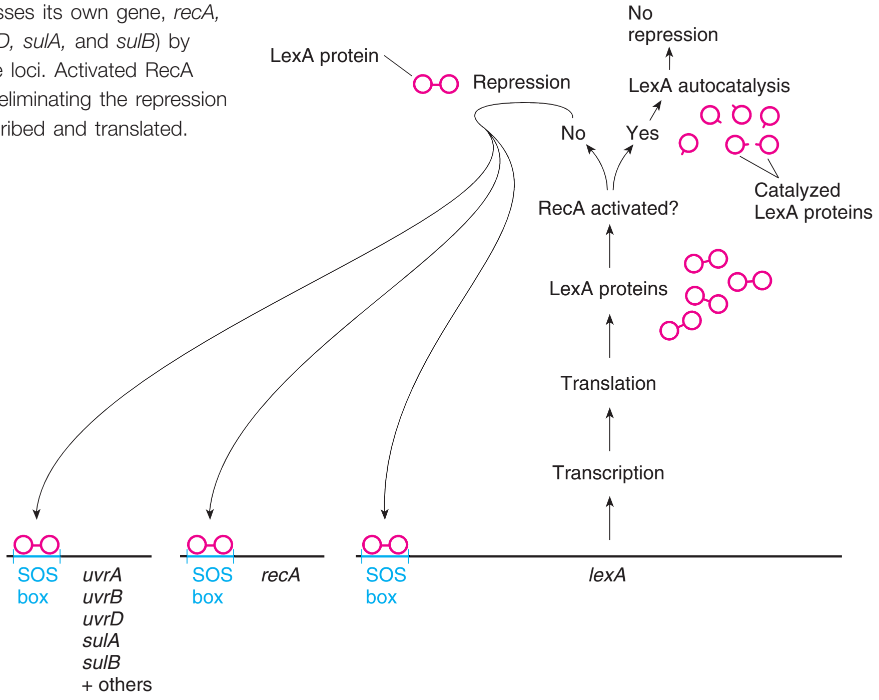


**Figure 12.37** The LexA protein represses its own gene, *recA,* and several other loci (*uvrA, uvrB, uvrD, sulA,* and *sulB*) by binding at the SOS box in each of the loci. Activated RecA protein causes autocatalysis of LexA, eliminating the repression of all these loci, which are then transcribed and translated.

## RECOMBINATION

Although recombination, the nonparental arrangement of alleles in progeny, can come about both by independent assortment and crossing over, we are concerned here with recombination due to crossing over between homologous pieces of DNA **(homologous recombination).** We briefly discuss transpositional recombination in chapter 14 and site-specific recombination (e.g., λ integration) in chapters 7, 14, and 16.

Recombination is a **breakage-and-reunion** process. Homologous parts of chromosomes come into apposition and are then reconnected in a crosswise fashion (see fig. 6.4). This general model fits what we know about the concordance of recombination and repair: Both involve breakage of the DNA and a small amount of repair synthesis, and both involve some of the same enzymes.

### *Double-Strand Break Model of Recombination*

In 1964, R. Holliday suggested a model of homologous recombination that involved simultaneous breaks in one strand each of the two double helices that were to cross over. In 1983, J. Szostak and colleagues put forth a different model, initiated by a double-strand break in one of the double helices. At first, this model was not considered seriously because a double-strand break was thought too dangerous a DNA lesion for cellular enzymes to create. However, we now know that the double-strand break model is generally correct, and we refer to the **Holliday junction** for an intermediate stage in the process. The model depends on DNA complementarity between the recombining molecules and is thus a model of great precision.

We begin with two double helices lined up as they would be, for example, in a meiotic tetrad, ready to undergo recombination (fig. 12.38*a*). The first step of the process is a double-stranded break in one of the double helices. In eukaryotes, the protein Spo11 accomplishes

R. Holliday (1932– ). (Courtesy of James L. German, III, M.D.)

**Table 12.5** Some of the Enzymes and Proteins Involved in DNA Repair in *E. coli,* Not Including DNA Polymerase I and III, DNA Ligase, and Single-Strand Binding Proteins

| Enzyme | Gene | Action |
|---|---|---|
| **Damage Reversal** | | |
| DNA photolyase | *phr* | Undimerizes thymine dimers |
| DNA methyltransferase | *ada* | Demethylates guanines in DNA |
| **Base Excision Repair** | | |
| Uracil-DNA glycosylase | *ung* | Removes uracils from DNA |
| Endonuclease IV | *nfo* | Nicks AP sites on the 5′ side |
| Exonuclease, lyase, or phosphodiesterase | *several* | Removes base-free nucleotide |
| **Nucleotide Excision Repair** | | |
| UvrA | *uvrA* | With UvrB, locates thymine dimers and other distortions |
| UvrB | *uvrB* | Nicks DNA on the 3′ side of the lesion |
| UvrC | *uvrC* | Nicks DNA on the 5′ side of the lesion |
| UvrD (helicase II) | *uvrD* | Unwinds oligonucleotide |
| **Mismatch Repair** | | |
| MutH | *mutH* | Nicks DNA at recognition sequence |
| MutL | *mutL* | Recognizes mismatch |
| MutS | *mutS* | Binds at mismatch |
| MutU (UvrD) | *mutU* | Unwinds oligonucleotide |
| Exonucleases | *recJ, xseA, sbcB* | Degrades unwound oligonucleotide |
| DNA methylase | *dam* | Methylates 5′-GATC-3′ DNA sequences |
| **Double-Strand Break Repair** | | |
| Ku | *Ku70, Ku80* | Binds to broken chromosomal ends |
| $PK_{CS}$ | $PK_{CS}$ | Protein kinase |
| DNA ligase IV | *LIG4* | Ligates broken ends of DNA |
| XRCC4 | *XRCC4* | Stabilization protein |
| **Postreplicative Repair** | | |
| Polymerase IV | *DinB* | DNA polymerase |
| Polymerase V | *UmuC, UmuD* | DNA polymerase |
| Polymerase η | *RAD30* | DNA polymerase |
| Polymerase ζ | *REV3, REV7* | DNA polymerase |
| RecA | *recA* | Single-stranded DNA invades double-stranded DNA; causes LexA to autocatalyze; protease |
| LexA | *lexA* | Represses SOS proteins |
| SulA, SulB | *sulA, sulB* | Inhibit cell division |

this. The break is followed by 5′ → 3′ exonuclease activity to widen the gaps formed in the double helix and create 3′ single-stranded tails (fig. 12.38*b, c*). These tails are coated with RecA protein that then catalyzes the invasion of one of the single strands into the intact double helix in direct apposition (fig. 12.38*d*). Repair of single-stranded DNA by DNA polymerase I and DNA ligase then replaces sections of previously digested DNA (fig. 12.38*e*). At this point, there is no “lost” genetic material; however the two double helices are interlocked and need to be freed of each other. Before that happens, however, **branch migration** can take place, a process in which the crossover point can slide down the duplexes (fig. 12.38*f*). In *E. coli,* the RuvAB complex, the product of the *ruvA* and *ruvB* genes that together form an ATP-dependent motor, moves the junction point

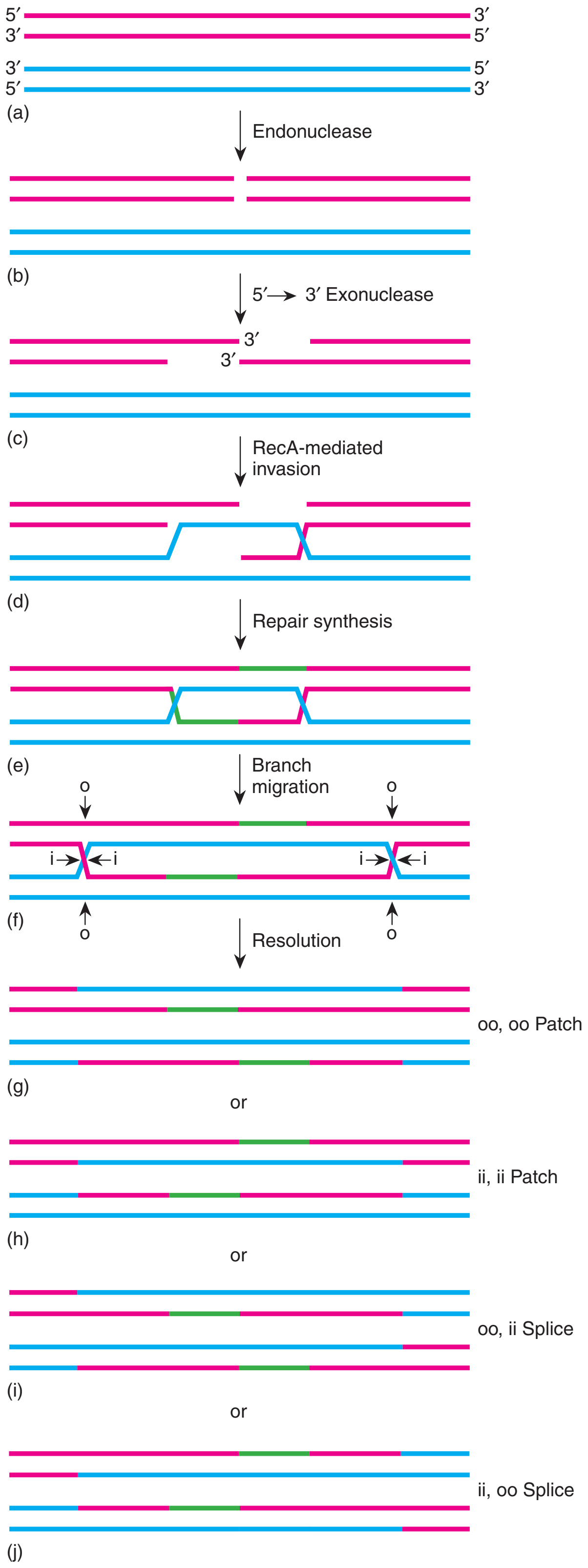


**Figure 12.38** The double-strand break model of genetic recombination. Two homologous duplexes (*a: red, blue*) of the four present in a meiotic tetrad are shown. An endonuclease creates a double-stranded break in one of the duplexes (*b*). A 5′→ 3′ exonuclease then digests away from the break in both directions, creating 3′ tails (*c*). RecA-mediated invasion occurs in the second duplex (*d*), followed by repair synthesis to close all gaps (*e*). Branch migration then takes place (*f*). Each of the Holliday junctions is then resolved independently, either by nicks in the two outer strands (*o*) or the two inner strands (*i*). Therefore, four resolution structures are possible (*g–j*). In patches, the ends of each duplex are the same as the original, indicating that there may not be recombination for loci flanking the point of crossover. In splices, the ends of each duplex have recombined, indicating that flanking loci may have crossed over. (Reprinted from *Cell*, Vol. 87, Frank Stahl, "Meiotic Recombination in Yeast: Coronation of the Double-Stranded-Break Repair Model," pp. 965–968, Copyright © 1996, with permission from Elsevier Science.)

(fig. 12.39). As the junction points move, they create heteroduplex DNA, places where the two strands of each double helix come from different original helices. These stretches have the potential to produce mismatches where the two chromatids differed originally. To resolve the cross-linked duplexes, a second cut at each junction is required.

Each of the two crossover points is a Holliday junction. If we open these junctions, we can see that each can be resolved in two different ways. (RuvC endonuclease, the protein product of the *ruvC* gene, resolves the Holliday junctions in *E. coli;* see fig. 12.39. RuvC cuts the Holliday junction at the consensus sequence 5′[A or T]TT[G or C]-3′. The cut is on the 3′ side of the two thymines.) Since there are two Holliday junctions per crossover, there are four potential combinations, as shown in figure 12.38*g–j.* Some of these combinations produce patches, where no recombination takes place among loci to the sides of the hybrid piece. Other combinations produce splices, where reciprocal recombination of loci takes place at the ends. The Holliday junctions can be seen in the electron microscope (fig. 12.40). Note that homology-directed recombination to repair double-strand breaks is similar to the process shown in figure 12.38.

## *Bacterial Recombination*

In bacterial recombination, a linear molecule recombines with a circular molecule (see fig. 7.15). Usually, invading DNA originates in the linear molecule. The RecBCD protein, whose subunits are the products of the *recB, recC,* and *recD* loci, initiates the first steps in forming an invading linear DNA molecule. RecBCD is a helicase, an exonuclease, and an endonuclease.

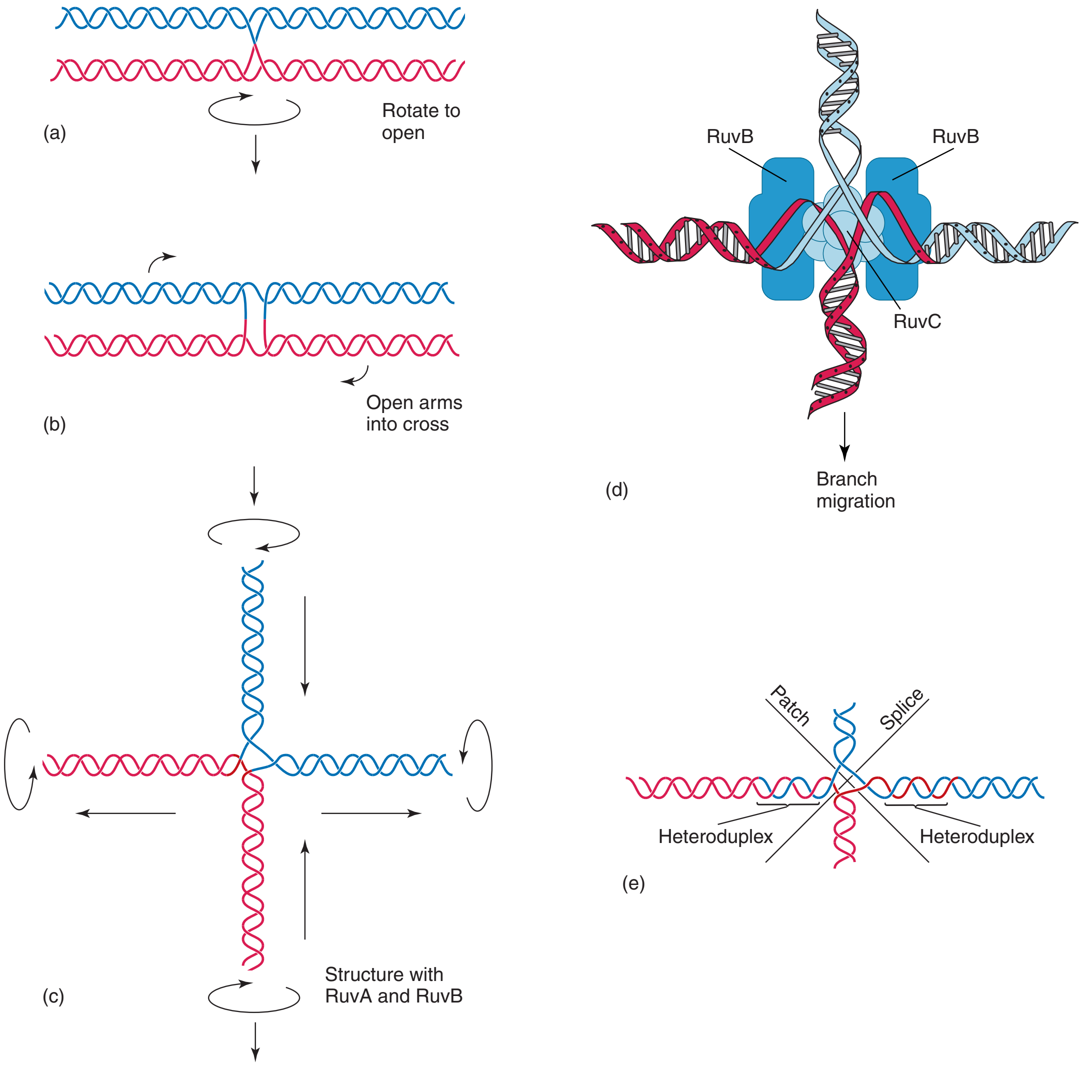


**Figure 12.39** Branch migration at a Holliday junction. (*a*) Two double helices (*red, blue*) are connected by a crossover. (*b*) The structure opens when one of the double helices rotates. (*c*) We further clarify the structure by separating the arms into an open cross, showing the direction in which the arms move during branch migration (*arrows:* the equivalent of pulling out on the left and right arms, drawing in the top and bottom arms). (*d*) A more realistic drawing with the RuvA and RuvB proteins, indicating one of the RuvA tetramers behind the center of the cross. A second tetramer (not shown) is located above the cross center, forming a RuvA sandwich of the cross center. The RuvB hexamers are shown on either side of the cross. (*e*) RuvC can resolve the cross to form either a splice or a patch, depending on which cut is made. (Reprinted with permission from *Nature,* Vol. 374, C. Parsons, et al., "Structure of a Multisubunit Complex that Promotes DNA Branch Migration." Copyright © 1995 Macmillan Magazines Limited.)

The RecBCD protein enters a DNA double helix from one end and travels along it in an ATP-dependent process. As it travels along the DNA, it acts as a 3′ → 5′ exonuclease, degrading one strand of the linear double helix (fig. 12.41). This process continues until RecBCD comes to a **chi site,** the sequence 5′-GCTGGTGG-3′, which appears about a thousand times on the *E. coli* chromosome. RecBCD's recognition of that sequence attenuates its 3′ → 5′ exonuclease activity and enhances its 5′ → 3′ exonuclease activity, begun after an endonucleolytic cleavage. From that point on, RecBCD creates a 3′ overhang or tail. That tail is coated by RecA and then

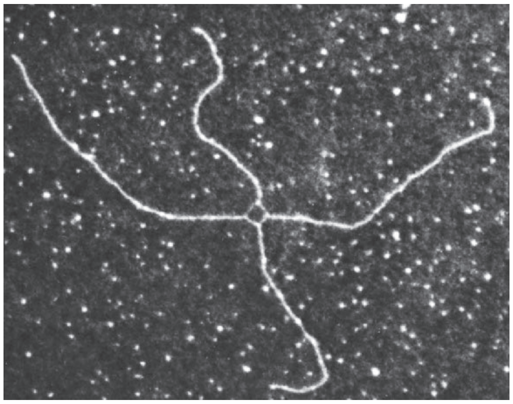


**Figure 12.40** A Holliday intermediate structure, equivalent to the structure seen in figure 12.39*c*. Each arm is about 1 micron long. (H. Potter and D. Dressler. "DNA recombination: In vivo and in vitro studies," *Cold Spring Harbor Symposium on Quantitative Biology,* Volume XLIII, 1979, pp. 969–85.)

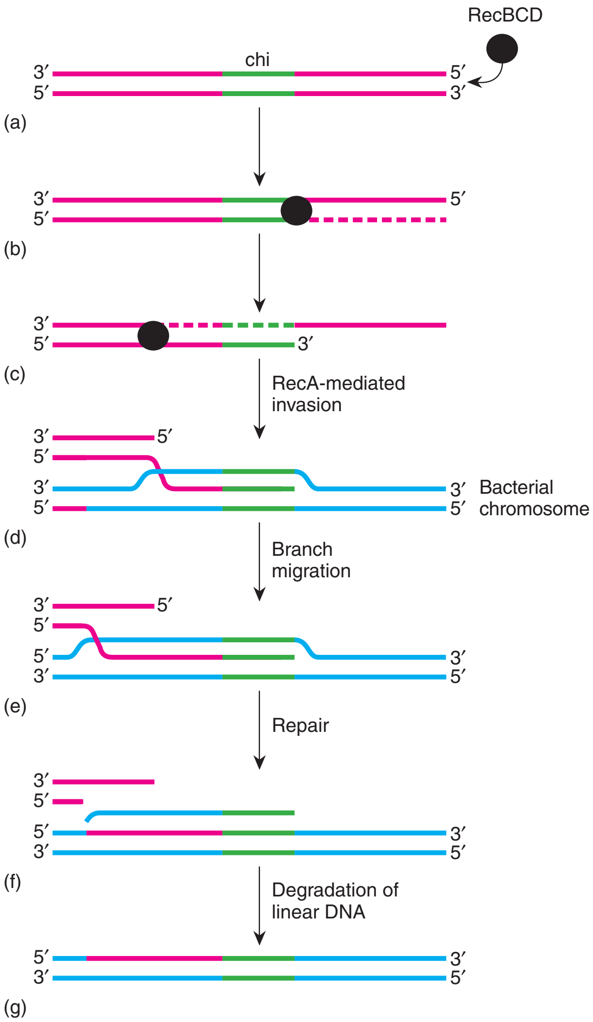


**Figure 12.41** RecBCD enters a linear DNA double helix (*red*) at one end and travels along it, digesting the 3′ strand. When the protein encounters a chi site (*green*), it cuts the other strand and begins acting as a 3′ → 5′ exonuclease, creating a 3′ overhang (*b, c*). The 3′ overhang can then invade a double helix mediated by RecA. Repair and degradation of the linear DNA results in hybrid DNA in the bacterial chromosome, which can be fixed by the mismatch repair system.

invades the circular bacterial chromosome to initiate a crossover event (fig. 12.41). After this pairing, the unpaired segments of the double helix of the bacteria and the exogenote are both degraded. Finally, DNA ligase seals the circular double helix. The resulting hybrid DNA will then be open to mismatch repair that can restore either original base pairs or base pairs from the invading DNA (see below).

## Hybrid DNA

The result of bacterial recombination or meiotic recombination with branch migration is a length of hybrid DNA. This **hybrid DNA,** also called **heterozygous DNA** or **heteroduplex DNA,** has one of two fates, if we assume a difference in base sequences in the two strands. Either the heteroduplex can separate unchanged at the next cell division, or the cell's mismatch repair system can repair it (fig. 12.42). Without appropriate methylation cues, the mismatch repair system can convert the CA base pair to either a CG or a TA base pair. If TA were the original bacterial base pair, conversion to CG would be a successful recombination, whereas return of the CA to TA would be restoration rather than recombination.

Recombination in yeast, or any other eukaryote, generates two heteroduplexes. The repair process can cause **gene conversion** (fig. 12.43), the alteration of progeny ratios indicating that one allele was converted to another, a phenomenon seen in up to 10% of yeast asci. The mismatched AC will be changed to an AT or a GC base pair; the mismatched TG base pair will be changed to TA or CG. The result of the repair, as shown at the bottom of figure 12.43, can be gene conversion in which an expected ratio of 2:2 ($a^- \ a^- \ a^+ \ a^+$) is converted to a 3:1 ratio ($a^- \ a^- \ a^- \ a^+$) or a 1:3 ratio ($a^- \ a^+ \ a^+ \ a^+$). If the heteroduplexes are not repaired, then a single cell generates both kinds of offspring after one round of DNA replication. Thus, the colony from the cell will be half wild-type ($a^+$) and half mutant ($a^-$). We see this phenomenon only in an Ascomycete fungus, such as yeast, in which all the products of a single meiosis remain together.

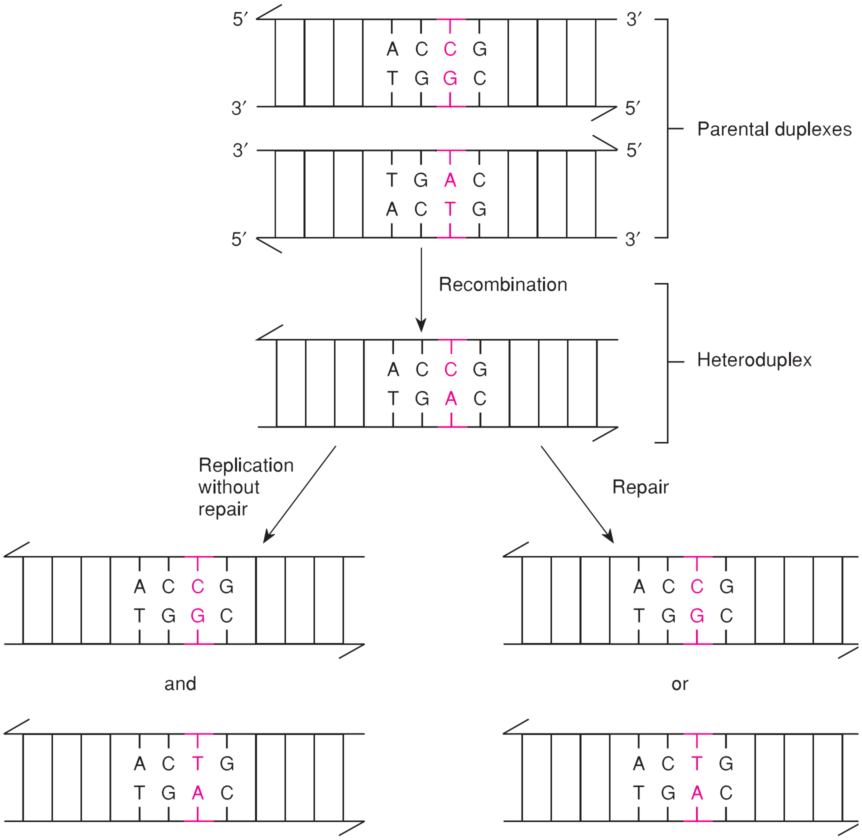


**Figure 12.42** Fate of a heteroduplex DNA. *Recombination* results in heteroduplex DNA with mismatched bases. *Replication without repair* produces two different daughter molecules. *Repair* converts the mismatched base pair to one or the other normal base pair.

## SUMMARY

**STUDY OBJECTIVE 1:** To look at the nature of mutation in prokaryotes 316–317

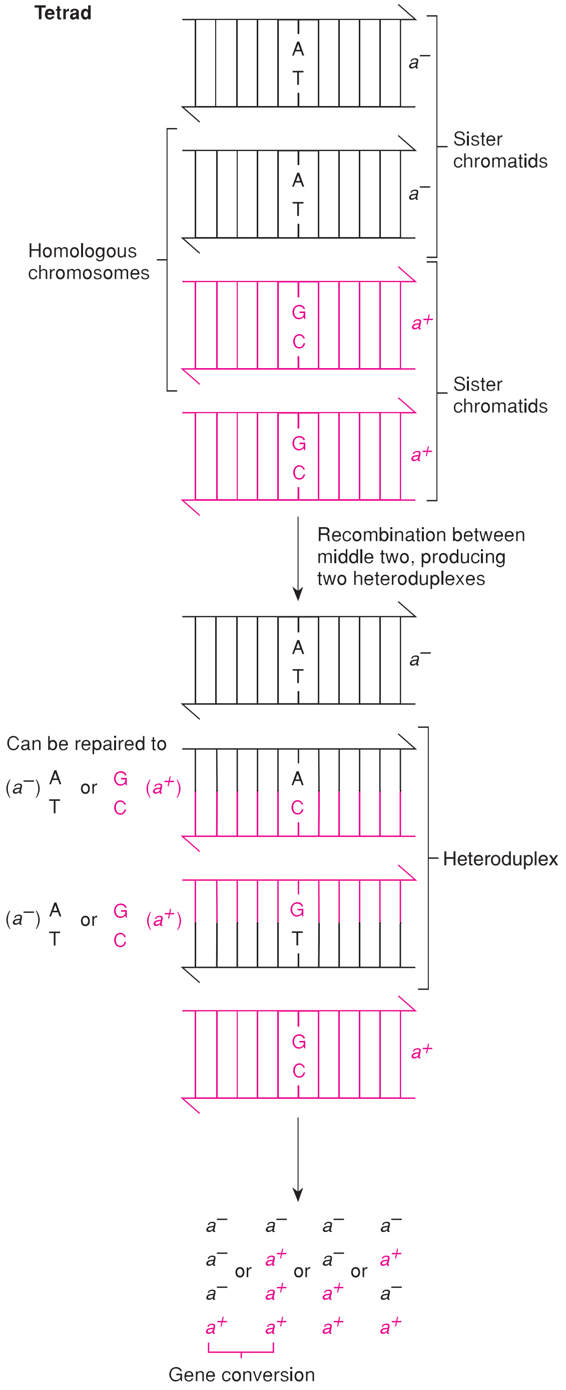
In 1943, Luria and Delbrück demonstrated that bacterial changes are true mutations similar to mutations in higher organisms. They showed that a high variability occurs in the number of mutants in small cultures as compared with the number of mutants in repeated subsamples of a large culture. Mutations thus occur spontaneously; they are caused by mutagens, which include chemicals and radiation. This chapter is concerned primarily with point mutations rather than changes in whole chromosomes or chromosomal parts.

**STUDY OBJECTIVE 2:** To analyze functional and structural allelism and examine the mapping of mutant sites within a gene 317–324

Allelism is defined by the *cis-trans* complementation test. Complementation implies independent loci, or nonallelic genes. Lack of complementation implies allelism. Functional alleles that differ from each other at the same nucleotides are also called structural alleles. Benzer used T4 phages to do fine-structure studies using complementation testing and deletion mapping.

**STUDY OBJECTIVE 3:** To verify the colinearity of gene and protein 324–325

Yanofsky demonstrated colinearity of the gene and protein. He had the advantage of working with a gene whose protein product was known.



Tetrad
A
T
$a^-$
A
T
$a^-$
Sister chromatids
Homologous chromosomes
G
C
$a^+$
Sister chromatids
G
C
$a^+$
Recombination between middle two, producing two heteroduplexes
A
T
$a^-$
Can be repaired to
($a^-$) A T or G C ($a^+$)
A
C
Heteroduplex
($a^-$) A T or G C ($a^+$)
G
T
G
C
$a^+$
$a^-$ $a^-$ $a^-$ $a^+$ or $a^-$ $a^+$ $a^+$ $a^+$ or $a^-$ $a^-$ $a^+$ $a^+$ or $a^-$ $a^+$ $a^-$ $a^+$
Gene conversion

**Figure 12.43** Recombination and repair can cause gene conversion. During recombination, heteroduplex DNA is formed, containing mismatched base pairs. Without methylation cues, repair enzymes convert the mismatch to a complementary base pair, in a random fashion—that is, an AC base pair can be converted to either an AT ($a^-$ allele) or a GC ($a^+$ allele) base pair. Two of the four possible repair choices create 3:1 ratios of alleles rather than the expected 2:2 ratios in the offspring. The 3:1 ratio represents gene conversion.

### STUDY OBJECTIVE 4: To study mutagenesis 325–338

After a mutation, the normal phenotype, or an approximation of it, can be restored either by back mutation or suppression. Intragenic suppression occurs when a second mutation within the same gene causes a return of normal or nearly normal function. Intergenic suppression occurs when a second mutation happens, usually in a transfer RNA gene, that counteracts the original mutation. Nonsense, missense, and frameshift mutations can all be suppressed.

Spontaneous mutation probably occurs primarily because of tautomerization of the bases of DNA. If a base is in the rare form during DNA replication, it can form unusual or mutant base pairings. We describe the mechanisms of action of the most common mutagens.

### STUDY OBJECTIVE 5: To investigate the processes of DNA repair and recombination 339–352

DNA repair processes can be divided into four categories: damage reversal, excision repair, double-strand break repair, and postreplicative repair. Photoreactivation is an example of damage reversal. Thymine dimers are undimerized by a photolyase enzyme in the presence of light energy. Excision repair removes a damaged section of a DNA strand. Repair enzymes fill in the gap. Excision repair can be divided into three types. In base excision repair, bases are removed by environmental causes or by glycosylases that sense damaged bases. AP endonuclease and an exonuclease, phosphodiesterase, or lyase then removes the base-free nucleotide. Some enzymes use base flipping to gain access to nucleotides in the double helix.

In nucleotide excision repair, enzymes in the Uvr system remove a patch containing the lesion, usually a thymine dimer. In mismatch repair, enzymes of the Mut system use methylation cues to remove a progeny patch containing the mismatch.

Double-strand break repair relies on one of two mechanisms. In nonhomologous end joining, the cell simply brings the broken ends back together. In homology-directed recombination, the cell repairs the broken ends using a recombinational mechanism.

Postreplicative repair fills in gaps left by DNA polymerase III. Some polymerase enzymes can use lesions as templates. Otherwise, the RecA protein is central to the process. A single strand from the undamaged duplex is used to fill the gap in the damaged duplex. Single-stranded DNA induces the SOS response, which temporarily eliminates LexA-mediated repression.

Recombination in eukaryotes begins with a double-stranded break in one double helix, followed by invasion of one of the ends into the other double helix. Repair, ligation, and branch migration follow. The crossover points, called Holliday junctions, need to be resolved, resulting in patches and splices. In *E. coli,* the RecBCD protein invades linear DNA, creating tails for invasion of the circular bacterial chromosome. Recombination results in heteroduplex DNA, which, if repaired, can lead to gene conversion. Thus, a battery of enzymes within the cell can modify DNA. These enzymes serve in DNA replication, repair, and recombination.

## SOLVED PROBLEMS

**PROBLEM 1:** An investigator isolates two recessive wing mutants of *Drosophila melanogaster.* The flies differ in wing vein pattern. Are the mutations that cause these variants allelic?

***Answer:*** To verify allelism, the investigator must create a heterozygote of the two mutations by either mating the flies, if they are of opposite sexes, or breeding each mutant into a separate stock for new matings. If the heterozygotes are of the wild-type, then the mutations are not allelic. If the heterozygote has a mutant phenotype, then we presume the mutations are functional alleles. (Allelism should be verified in females to be sure that the locus is not on the X chromosome, since males have only one.) If the mutations are functional alleles, it is possible to determine whether they are also structural alleles by looking for wild-type offspring of the heterozygote. If they occur at a rate higher than the background mutation rate, the alleles are not structural alleles. If wild-type offspring occur at the mutation rate, the alleles are presumably structural.

**PROBLEM 2:** What is the difference between mismatch repair and AP repair?

***Answer:*** Both processes are similar in that they entail removal of an incorrect base in a DNA double helix by an excision process followed by a repair process. The processes differ in the event that triggers them. Mismatch repair is triggered by a base pair that does not occupy the correct space in the double helix—that is, by a non-Watson and Crick pairing (not AT or GC). AP repair is triggered by enzymes that recognize a missing base.

**PROBLEM 3:** What role does the RecBCD protein play in recombination?

***Answer:*** For recombination to take place in *E. coli,* a single strand of DNA from the exogenote must insinuate itself into the chromosomal double helix with the help of the RecA protein. It is the RecBCD protein that creates the single-stranded DNA. It does so by traveling down the double helix, creating a single-strand tail in its wake. At chi sites, it switches the activity of the enzyme from a $3' \rightarrow 5'$ exonuclease to a $5' \rightarrow 3'$ exonuclease, creating a $3'$ tail. RecA can then act on this single-stranded tail to initiate recombination.

## EXERCISES AND PROBLEMS[*]

**MUTATION**

1. Construct a data set that Luria and Delbrück might have obtained that would prove the mutation theory wrong.
2. What types of enzymatic functions are best studied using temperature-sensitive mutations?
3. Seven arginine-requiring mutants of *E. coli* were independently isolated. All pairwise matings were done (by transduction) to determine the number of loci (complementation groups) involved. If a (+) in the following figure indicates growth and a (−) no growth on minimal medium, how many complementation groups are involved? Why is only “half” a table given? Must the upper left to lower right diagonal be all (−)?

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | – | + | + | + | + | – | – |
| 2 |  | – | + | + | – | + | + |
| 3 |  |  | – | – | + | + | + |
| 4 |  |  |  | – | + | + | + |
| 5 |  |  |  |  | – | + | + |
| 6 |  |  |  |  |  | – | – |
| 7 |  |  |  |  |  |  | – |

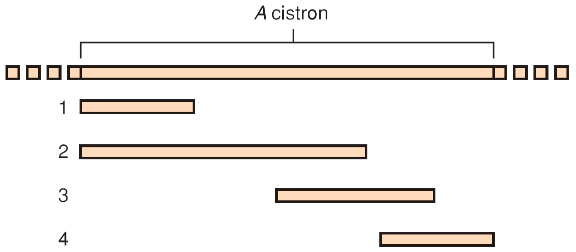
4. Several *r*II mutations (M to S) have been localized to the *A* cistron because of their failure to complement with a known deletion in that cistron. The phages carrying these mutations are then mated pairwise with the following series of subregion deletions. The mating is done on *E. coli* B and plated out on *E. coli*

---

[*] Answers to selected exercises and problems are on page A-13.

K12 (see the figure). A (+) shows the presence of plaques on K12, whereas a (−) shows an absence of growth. Look at the deletion map of the area and localize each of the *r*II *A* mutations on this map.

| | Deletion | | | |
|---|---|---|---|---|
| **Mutant** | **1** | **2** | **3** | **4** |
| **M** | + | + | − | + |
| **N** | + | − | − | + |
| **O** | + | + | + | − |
| **P** | + | + | − | − |
| **Q** | + | − | + | + |
| **R** | − | − | + | + |
| **S** | − | − | − | + |



Relative positions of deletions

5. A *Drosophila* worker isolates four eye-color forms of the fly: wild-type, white, carmine, and ruby. (The worker does not know that white, carmine, and ruby are three separate loci on the X chromosome.) What crosses could the researcher make to determine allelic relations of the genes? What results would be expected? A new mutant, eosin, is isolated. What crosses should be carried out to determine that eosin is an allele of white?
6. Define structural and functional alleles. What is the *cis* part of a *cis-trans* complementation test?
7. Did Benzer and Yanofsky work with genes that had intervening sequences? What relevance might introns have to their work?
8. How can intra-allelic complementation result in incorrect conclusions about allelism?
9. *E. coli* bacteria of strain K12 are lysogenic for phage $\lambda$. Why won't *r*II mutants of phage T4 grow in these bacteria?
10. Diagram the tautomeric base pairings in DNA. What base pair replacements occur because of the shifts?
11. What is the difference between a substrate and a template transition mutation?
12. Describe two mechanisms for transversion mutagenesis.
13. 5-bromouracil, 2-aminopurine, proflavin, ethyl ethane sulfonate, and nitrous acid are chemical mutagens. What does each do?
14. A point mutation occurs in a particular gene. Describe the types of mutational events that can restore a functional protein, including intergenic events. Consider missense, nonsense, and frameshift mutations.
15. Why does misalignment result in addition or deletion of bases?
16. What are the differences and similarities between intergenic and intragenic suppression?
17. Eight independent mutants of *E. coli* requiring tryptophan (*trp*) are isolated. Complementation tests are performed on all pairwise combinations. Based on the results shown, determine how many genes you have identified and which mutants are in which genes (+ = complementation; − = no complementation).

| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
|---|---|---|---|---|---|---|---|---|
| **1** | − | + | − | − | + | + | + | − |
| **2** | | − | + | + | − | + | + | + |
| **3** | | | − | − | + | − | − | − |
| **4** | | | | − | + | + | + | − |
| **5** | | | | | − | + | + | + |
| **6** | | | | | | − | − | + |
| **7** | | | | | | | − | + |
| **8** | | | | | | | | − |

18. Complementation tests are usually done with recessive mutations for if the mutations were dominant, all progeny would be mutant, regardless of whether the genes are allelic. Suppose you have isolated in a diploid species two independent dominant mutations that each confer resistance to the drug cycloheximide. Call these mutations *Chx-1* and *Chx-2.* What crosses can you perform to determine whether the mutations are allelic? Your crosses should allow you to determine whether the mutations are allelic, nonallelic and unlinked, or nonallelic and linked.
19. A series of overlapping deletions in phage T4 are isolated. All pairwise crosses are performed, and the progeny scored for wild-type recombinants. In the following table, + = wild-type progeny recovered; − = no wild-type progeny recovered.

| | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|
| **1** | − | + | − | − | − |
| **2** | | − | + | + | − |
| **3** | | | − | + | + |
| **4** | | | | − | − |
| **5** | | | | | − |

**a.** Draw a deletion map of these mutations.

**b.** A point mutation, 6, is isolated and crossed with all of the deletion strains. Wild-type recombinants are recovered only with strains 2 and 3. What is the location of the point mutation?

**20.** Hydroxylamine is a chemical that causes exclusively C → T transition mutations. Can hydroxylamine reverse nonsense mutations? Explain.

**21.** A nonsense suppressor is isolated and shown to involve a tyrosine transfer RNA. When this mutant transfer RNA is sequenced, the anticodon turns out to be normal, but a mutation is found in the dihydrouridine loop. What does this finding suggest about how a transfer RNA interacts with the messenger RNA?

**22.** Devise selection-enrichment procedures for isolating the following kinds of mutants:

**a.** extra-large bacterial cells

**b.** nonmotile ciliated protozoans

**23.** Two chemically induced mutants, *x* and *y*, are treated with the following mutagens to see if revertants can be produced: 2-amino purine (2AP), 5-bromouracil (5BU), acridine dye (AC), hydroxylamine (HA), and ethylmethanesulfonate (EMS). In the following table, + = revertants and − = no revertants. For each mutation, determine the probable base change that occurred to change the wild-type to the mutant.

| | Chemical | | | | |
|---|---|---|---|---|---|
| **Mutant** | **2AP** | **5BU** | **AC** | **HA** | **EMS** |
| *x* | − | + | − | + | + |
| *y* | + | − | − | − | − |

**24.** What situation will lead to a false positive in a complementation test—or, in other words, indicate two genes when, in fact, the mutations are in the same gene?

**25.** What situation will lead to false negatives in a complementation test—or, in other words, indicate mutations are in the same gene when, in fact, they are in different genes?

**26.** Suppose you repeat the Luria-Delbrück fluctuation test, but this time you look for *lac* colonies. Your "individual cultures" produce the following numbers of lac colonies: 20, 25, 22, 18, 24, 19, 17, 25, 26, and 18. Subsamples from the bulk culture give results identical to these. What can you conclude?

**27.** You have isolated a new histidine auxotroph, and, despite all efforts, you cannot produce any revertants. What probably happened to produce the original mutant?

### DNA REPAIR

**28.** UV light causes thymine dimerization. Describe the mechanisms, in order of efficiency, that can repair the damage. Name the enzymes involved.

**29.** What types of damage do excision repair endonucleases recognize?

**30.** What are the functions of the RecA protein? How is it involved in phage λ induction? (*See also* RECOMBINATION)

### RECOMBINATION

**31.** Diagram, in careful detail, a recombination by way of the double-strand break model. What enzymes are required at each step?

**32.** What are the different enzymes involved in reciprocal and nonreciprocal recombination?

## CRITICAL THINKING QUESTIONS

**1.** Charles Yanofsky demonstrated colinearity of the gene and its protein product. What are the alternatives to colinearity?

**2.** Comment on the statement that DNA is a molecule designed for replication and repair.

*Suggested Readings for chapter 12 are on page B-10.*

# 13

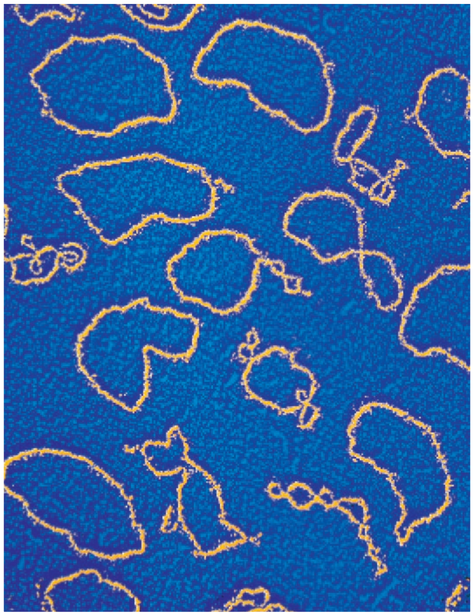
# GENOMICS, BIOTECHNOLOGY, AND RECOMBINANT DNA

## STUDY OBJECTIVES

1. To look at the techniques of gene cloning 358
2. To examine the techniques of creating restriction maps 377
3. To study the methods of DNA sequencing 383
4. To look at the goals and methods of the Human Genome Project 390
5. To look at the practical benefits and human issues of genetic engineering 397

## STUDY OUTLINE

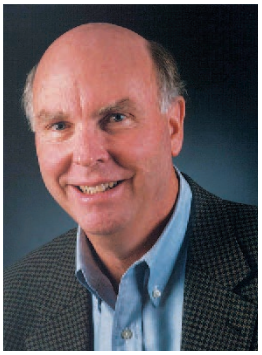
**Genomic Tools** 359
  Restriction Endonucleases 359
  Prokaryotic Vectors 360
  Cloning a Particular Gene 364
  Southern Blotting 366
  Probing for a Cloned Gene 368
  Eukaryotic Vectors 369
  Expression of Foreign DNA in Eukaryotic Cells 372
**Restriction Mapping** 377
  Constructing a Restriction Map 379
  Double Digests 379
  Restriction Fragment Length Polymorphisms 380
**Polymerase Chain Reaction** 381
**DNA Sequencing** 383
  The Dideoxy Method 383
  Creating a General-Purpose Primer 386
**Mapping and Sequencing the Human Genome** 390
  Locating a Gene of Interest 390
  The Human Genome Project 391
  Ethics 397
**Practical Benefits from Gene Cloning** 397
  Medicine 397
  Agriculture 398
  Industry 398
*Summary* 398
*Solved Problems* 399
*Exercises and Problems* 400
*Critical Thinking Questions* 404
**Box 13.1 The Recombinant DNA Dispute** 370
**Box 13.2 Cloning Dolly** 374
**Box 13.3 Genes Within Genes** 388



Artificially colored transmission electron micrograph of DNA plasmids from the bacterium *Escherichia coli.* These plasmids are used in genetic engineering.

(© Dr. Gopal Murti/SPL/Photo Rsearchers, Inc.)

In the spring of 2000, J. Craig Venter, CEO of Celera Genomics, and Francis Collins, director of the National Institutes of Health's Human Genome Research Institute, jointly announced that they and their colleagues had completed the sequence of the human genome. Although there is still work ahead to finish the project, the accomplishment was enormous. To some, it was working out the very secret of life. This accomplishment firmly established the science of **genomics,** the study of the mapping and sequencing of genomes.

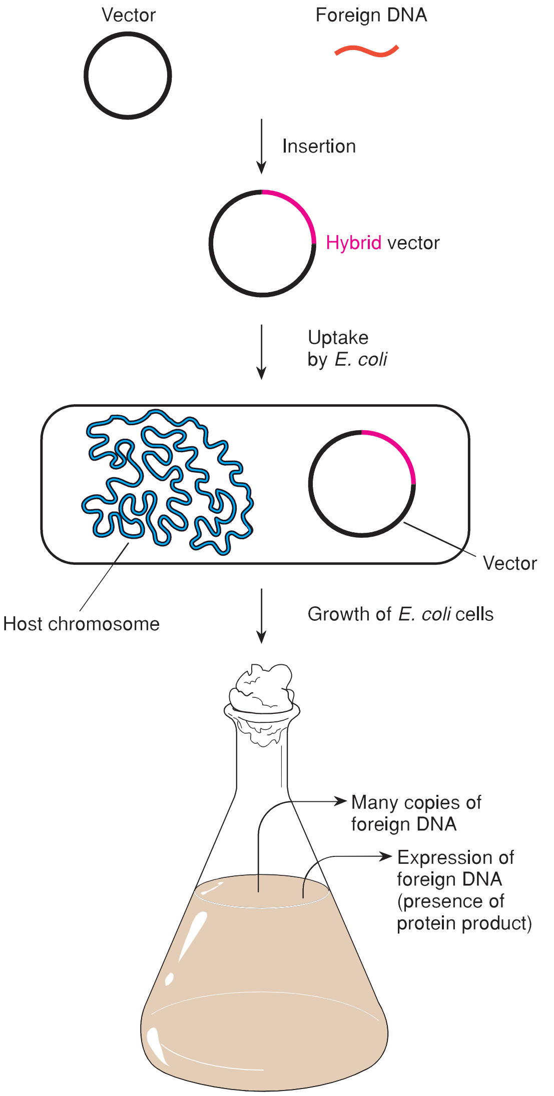


J. Craig Venter (1946– ).
(Courtesy of Celera Genomics.)

Francis S. Collins (1950– ).
(Courtesy of Francis Collins.)

Since the mid-1970s, the field of molecular genetics has undergone explosive growth, noticeable not only to geneticists, but also to medical practitioners and researchers, agronomists, animal scientists, venture capitalists, and the public in general. Medical practitioners and researchers have new treatments for diseases available. Agronomists see the possibility of greatly improved crop yields, and animal scientists have gained the possibility of greatly improving food production from domesticated animals. Geneticists and molecular biologists are gaining major new insights into understanding gene expression and its control.

The new DNA manipulation techniques, centered on the isolation, amplification, sequencing, and expression of genes, are based on the insertion of a particular piece of foreign DNA into a vector—a plasmid or phage. A plasmid is placed into a host cell, either prokaryotic or eukaryotic, which then divides repeatedly, producing numerous copies of the vector with its foreign piece of DNA. A phage simply multiplies in host cells (fig. 13.1). In both cases, the foreign piece of DNA is amplified in number; it can be expressed (transcribed and translated into a protein) when in a plasmid in a host cell. A commonly used host cell is *E. coli.* Following its amplification, the foreign DNA can be purified and its nucleotide sequence determined. When it is expressed, large quantities of the gene product of the foreign DNA can be obtained. The new technology is variously referred to as **gene cloning, recombinant DNA technology,** or **genetic engineering.** In this chapter, we look in detail at the methods and procedures of recombinant DNA technology, including DNA sequencing.



Vector
Foreign DNA
Insertion
Hybrid vector
Uptake by *E. coli*
Vector
Host chromosome
Growth of *E. coli* cells
Many copies of foreign DNA
Expression of foreign DNA (presence of protein product)

**Figure 13.1** Overview of recombinant DNA techniques. A hybrid vector is created, containing an insert of foreign DNA. The vector is then inserted into a host organism. Replication of the host results in many copies of the foreign DNA and, if the gene is expressed, quantities of the gene product. (All DNA shown is double-stranded.)

# GENOMIC TOOLS

## *Restriction Endonucleases*

In 1978, Nobel prizes in physiology and medicine were awarded to W. Arber, H. Smith, and D. Nathans for their pioneering work in the study of **restriction endonucleases.** These are enzymes that bacteria use to destroy foreign DNA, presumably, the DNA of invading viruses. The enzymes recognize certain nucleotide sequences **(restriction sites)** found on foreign DNA, usually from four to eight base-pairs long, and then cleave that DNA at or near those sites. (Restriction endonucleases were originally so named because they restricted phage infection among strains of bacteria. Phages that could survive in one strain could not survive in other strains with different restriction enzymes.)

Three types of restriction endonucleases are known. Their groupings are based on the types of sequences they recognize, the nature of the cut made in the DNA, and the enzyme structure. Types I and III restriction endonucleases are not useful for gene cloning because they cleave DNA at sites other than the recognition sites and thus cause random, unpredictable cleavage patterns. Type II endonucleases, however, cleave at the specific sites they recognize, leading to predictable cleavage patterns. The sites type II endonucleases recognize are inverted repeats; they have twofold symmetry. To see the symmetry, you must read outward from a central axis on opposite strands of the DNA. For example, the type II restriction endonuclease *Bam*HI recognizes

5′-GGA | TCC-3′
3′-CCT | AGG-5′

Reading out from the center (vertical line) is AGG on the top strand and AGG on the bottom strand. The sequence is, in a sense, **a palindrome,** a sequence that reads the same from either direction. (Palindrome is from the Greek *palindromos,* which means "to run back." The name Hannah and the numerical sequence 1238321 are palindromes.) In figure 13.2 are some palindromic sequences that type II restriction endonucleases recognize; well over one hundred type II enzymes are known.

The host cell protects its DNA not by being free of these restriction sites, but usually by methylating its DNA in these regions (fig. 13.3). The same sequences that the

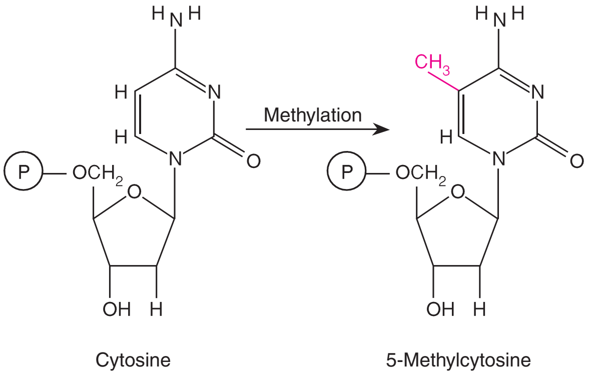

H N H
H
N
H
N
O
P
OCH$_2$
O
OH H
Methylation
CH$_3$
Cytosine
5-Methylcytosine

**Figure 13.3** A methylase enzyme adds a methyl group to cytosine, converting it to 5-methylcytosine.

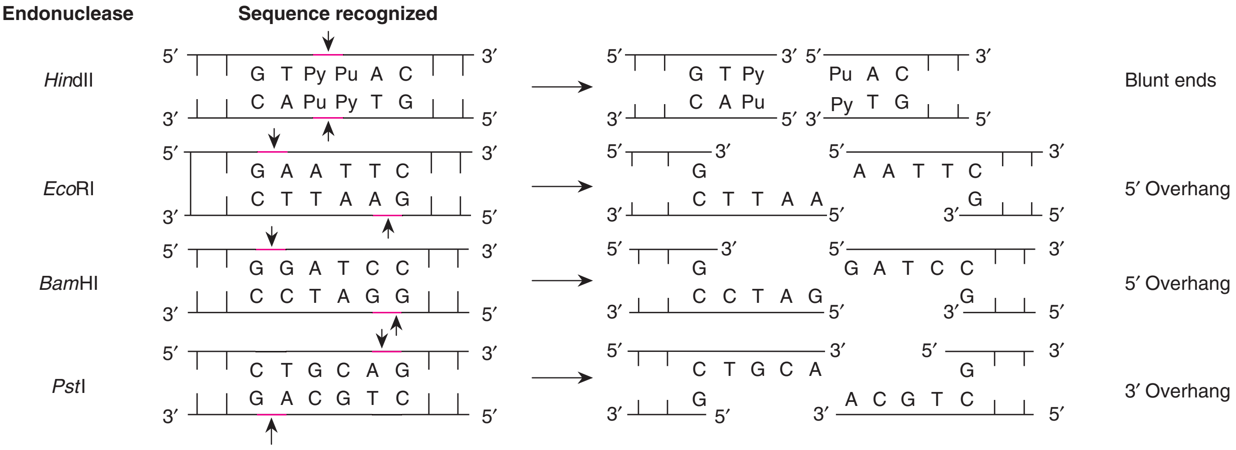

Endonuclease
Sequence recognized
*Hin*dII
5′ G T Py Pu A C 3′
3′ C A Pu Py T G 5′
5′ G T Py 3′ 5′ Pu A C 3′
3′ C A Pu 5′ 3′ Py T G 5′
Blunt ends
*Eco*RI
5′ G A A T T C 3′
3′ C T T A A G 5′
5′ G 3′ 5′ A A T T C 3′
3′ C T T A A 5′ 3′ G 5′
5′ Overhang
*Bam*HI
5′ G G A T C C 3′
3′ C C T A G G 5′
5′ G 3′ 5′ G A T C C 3′
3′ C C T A G 5′ 3′ G 5′
5′ Overhang
*Pst*I
5′ C T G C A G 3′
3′ G A C G T C 5′
5′ C T G C A 3′ 5′ G 3′
3′ G 5′ 3′ A C G T C 5′
3′ Overhang

**Figure 13.2** Sequences cleaved by various type II restriction endonucleases. Py is any pyrimidine and Pu is any purine. *Arrows* denote places where endonucleases cleave the DNA. In 1971, K. Danna and D. Nathans showed that a restriction endonuclease would consistently cut DNA into pieces of the same size. This precision and repeatability of enzyme action made enzymes useful for further research. Not all restriction endonucleases make staggered cuts with 3′ and 5′ overhangs; some produce blunt ends.

endonucleases attack in the unmethylated condition are protected when methylated. After host DNA replication, new double helices are hemimethylated; that is, the old strand is methylated but the new one is not. In this configuration, the new strand is quickly methylated (fig. 13.4). Foreign DNA, without methyl groups on either strand, is not methylated.

Restriction endonucleases are named after the bacteria from which they were isolated: *Bam*HI from *Bacillus amyloliquefaciens,* strain H; *Eco*RI from *E. coli,* strain RY13; *Hin*dII from *Haemophilus influenzae,* strain Rd; and *Bgl*I from *Bacillus globigii.* From here on, we will refer to type II restriction endonucleases simply as restriction enzymes.

Restriction enzymes cut the DNA in two different ways. For example, *Hin*dII cuts the recognition sequence down the middle, leaving "blunt" ends on the DNA (see fig. 13.2). We will discuss how pieces of DNA with blunt ends can be used in cloning. The staggered cuts made, for example, by *Bam*HI leave "sticky" ends (a 5′ overhang) that can reanneal spontaneously as hydrogen bonds form between the complementary bases (see fig. 13.2). The ability to reanneal these sticky ends, first demonstrated by S. Cohen, H. Boyer, and colleagues in 1971, opened up the field of gene cloning.

## Prokaryotic Vectors

With current technology it is routine to join together, in vitro, DNAs from widely different sources. In figure 13.5, we see how a circular DNA molecule cleaved by a specific restriction enzyme can recircularize if it is cleaved in only one place, or how different molecules with the same free ends can anneal to form hybrid molecules. Only the action of a DNA ligase is needed to make the molecules complete (see chapter 9).

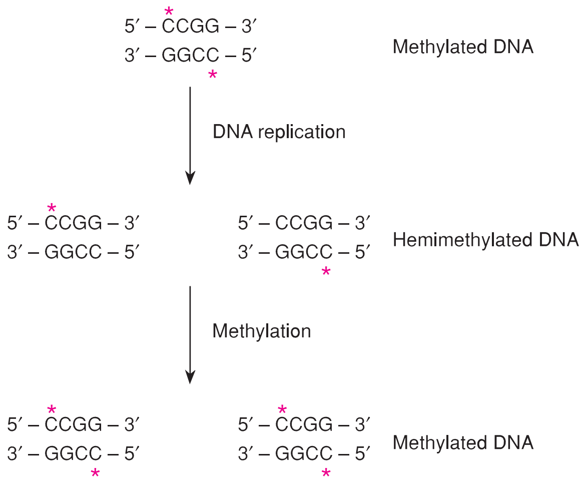


**Figure 13.4** Host DNA is methylated in the *Hpa*II restriction site. *Asterisks* indicate methyl groups on cytosines. After DNA replication, the DNA is hemimethylated; the new strands have no methyl groups. Hemimethylated DNA is then fully methylated by cellular enzymes.

One of the pieces of DNA involved in the annealing can be a plasmid, a piece of DNA that can replicate in a cell independently of the cellular chromosome. The **recombinant plasmid** (fig. 13.6) can be transferred into a cell. (A recombinant plasmid is also known as a **hybrid plasmid, hybrid vehicle, hybrid vector,** or **chimeric plasmid.** The latter is after the *chimera,* a mythological monster with a lion's head, a goat's body, and a serpent's tail.) Many procedures exist that can introduce this recombinant plasmid into a host cell. For example, a bacterial cell can be made permeable to this, or any, plasmid by the addition of a dilute solution of calcium chloride. Once inside the cell, the foreign DNA is replicated each time the plasmid DNA replicates.

Note that in the process of inserting a piece of foreign DNA, the restriction site is duplicated, with one copy at either end of the insert. This property makes it easy to remove the cloned insert at some future time, if needed, since restriction sites enclose it (fig. 13.6).

### Cloning with Restriction Enzymes

A few conditions must be met in order to succeed in cloning DNAs from different sources. A plasmid vehicle should be cleaved at only one point by the endonuclease. If it is cleaved at more than one point, it will fragment during the experiment. However, some phage vehicles must be cleaved at two points so that the foreign DNA can replace a length of the phage DNA rather than simply being inserted. Common vehicles, derivatives of phage λ, have been named **Charon phages** (pronounced "karon") after the mythical boatman of the River Styx. (See chapter 14 for a detailed discussion of phage λ.)

During normal phage infection (see chapter 7), only DNA the size of a phage genome is packaged into λ heads. Thus, for λ to be a useful vector, the foreign DNA must replace part of its DNA. We note that λ can function quite well as a hybrid vehicle with a 15,000 base-pair (15 kilobases, or 15 kb) section replaced by foreign DNA because that section of phage DNA is used for integration into the *E. coli* chromosome, a nonessential phage function. That is, the phage can infect a bacterium, replicate inside the bacterium, and burst out without the integration region. Genetic engineers have created a λ DNA molecule with the nonessential region missing and an *Eco*RI cleavage site in its place. Only hybrid DNA can thus be incorporated

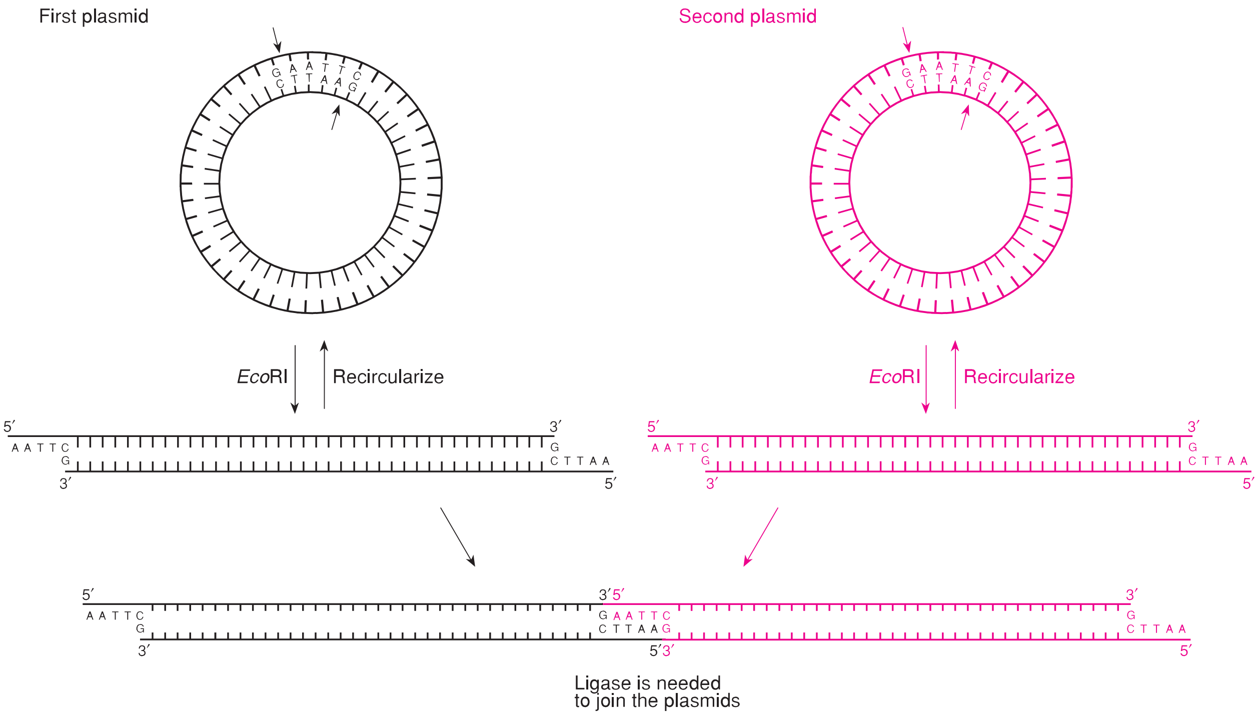


**Figure 13.5** Circular plasmid DNA with a palindrome recognized by *Eco*RI. After the DNA is cleaved by the endonuclease, it has two exposed ends that can join to recircularize the molecule or unite two or more linear molecules of DNA cleaved by the same restriction endonuclease. The final nicks are closed with DNA ligase. S. Cohen, H. Boyer, and their colleagues first joined plasmids with this technique in 1971.

into phage heads because the diminished phage DNA, without an insert, is too small to be properly packaged.

One disadvantage of cloning with normal *E. coli* plasmids is that they are unstable if the foreign DNA is very large, greater than about 15 kb. That is, if a large chromosomal segment is cloned, the plasmid tends to lose parts of the clone as the plasmid replicates. Primarily for this reason, geneticists began using phage λ as a vector (see fig. 7.21) because these phages could successfully maintain foreign DNA as large as 24 kb.

The phage chromosome is about 50 kb of DNA; within the phage head it is linear, and within the cell it is circular. The DNA to fill the phage head is recognized during infection because it has a small segment of single-stranded DNA called a *cos* site (twelve bases; derived from the term "cohesive ends") at either end. Reannealing the *cos* sites allows λ chromosomes to circularize when they enter a host cell; cutting the DNA at the *cos* site opens the circle into a linear molecule (fig. 13.7). Geneticists have taken advantage of these *cos* sites to clone even larger segments of foreign DNA because it turns out that even 24 kb is not adequate to study some eukaryotic genes or gene groups. Many eukaryotic genes are very large because of their introns and transcriptional control segments. DNA up to 50 kb can be cloned if *cos* segments are attached to either end with a plasmid origin of DNA replication and a selectable antibiotic gene. These *cos*-site-containing plasmids are called **cosmids** (fig. 13.7). Cosmids not only allow the cloning of very large pieces of DNA, they actually select for large segments of foreign DNA because small cosmids are not incorporated into phage heads. Thus, foreign DNA ranging from 2.5 to 50 kb in size can be cloned using plasmids, Charon phages, or cosmids. (Much larger pieces of DNA, about a million bases, can be cloned in yeast, as we will describe later.)

### *Selecting for Hybrid Vectors*

In the methods we have described, restriction enzymes separately cut both vector and foreign DNA. The two are then mixed in the presence of ligase. The many products that are created can be divided generally into three categories: vectors with foreign DNA, vectors without foreign DNA, and fragments. In a later section, we will discuss methods of finding a particular piece of foreign DNA in a vector. Here, we point out how vectors with inserts of any kind are selected.

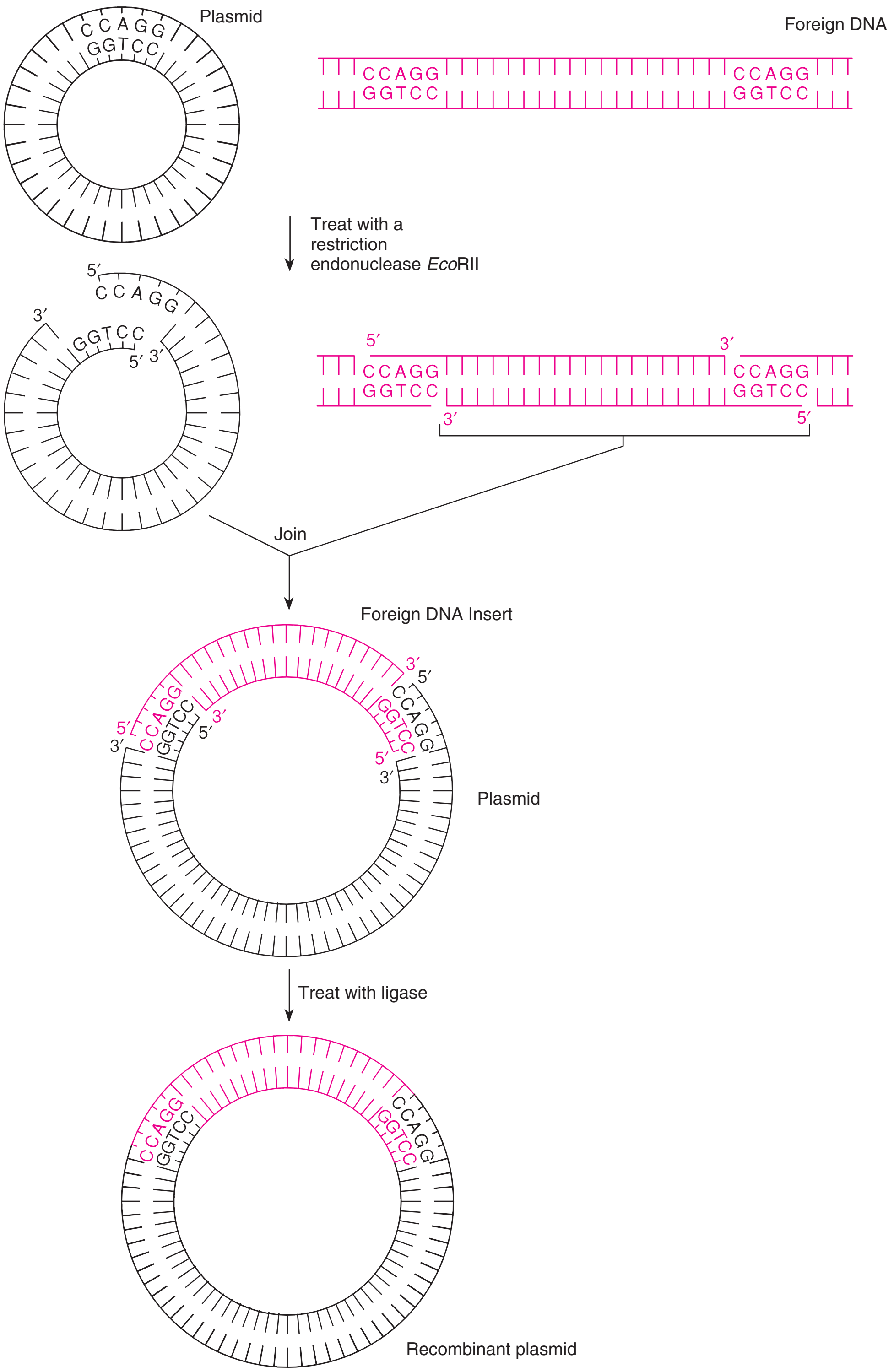

Plasmid
CCAGG
GGTCC
Foreign DNA
CCAGG
GGTCC
CCAGG
GGTCC
Treat with a restriction endonuclease *Eco*RII
5′
CCAGG
3′
GGTCC
5′ 3′
5′
3′
CCAGG
GGTCC
CCAGG
GGTCC
3′
5′
Join
Foreign DNA Insert
3′ 5′
CCAGG
GGTCC
5′ 3′
5′ 3′
CCAGG
GGTCC
3′ 5′
Plasmid
Treat with ligase
CCAGG
GGTCC
CCAGG
GGTCC
Recombinant plasmid

**Figure 13.6** Formation of a recombinant plasmid. The same restriction endonuclease, in this case *Eco*RII, is used to cleave both host and foreign DNA. Some of the time, cleaved ends will come together to form a plasmid with an insert of the foreign DNA. Ligase seals the nicks. P. Berg was the first scientist to clone a piece of foreign DNA when he inserted the genome of the SV40 virus into phage λ in 1973.

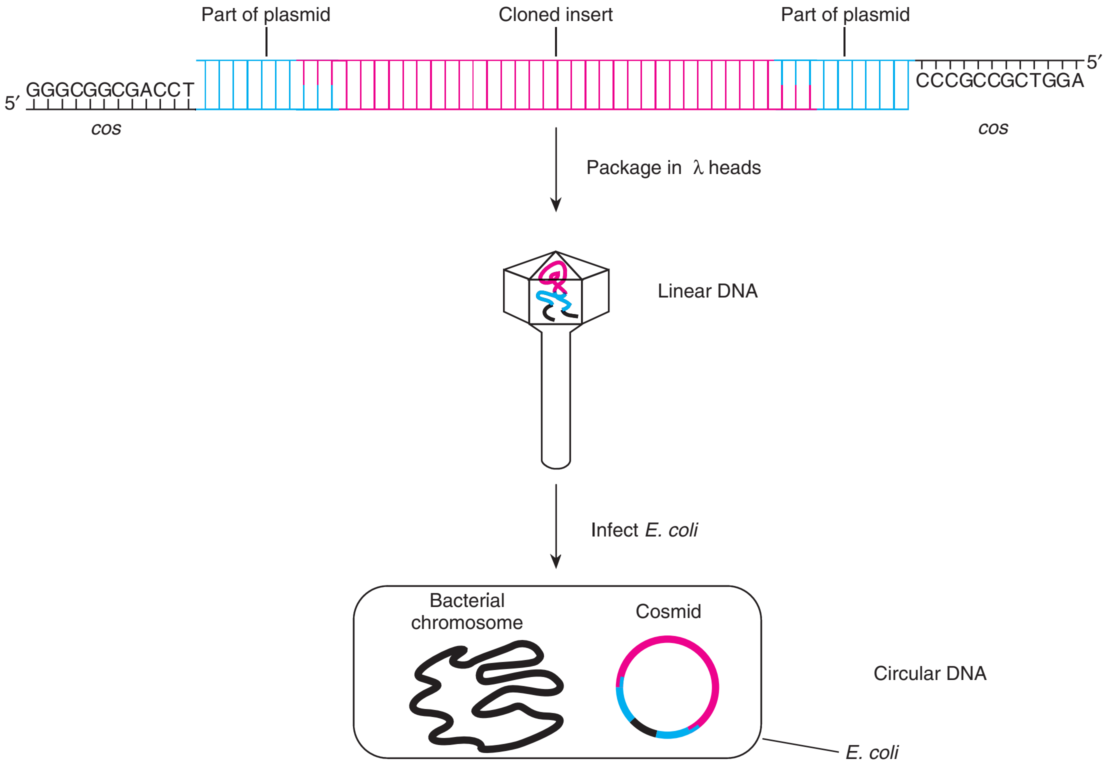


**Figure 13.7** A cosmid is a plasmid with *cos* sites that can be transferred into bacteria within phage lambda heads, a very efficient method of infection. The *cos* sites are single-stranded; they reanneal to a *circle* when inside the host. (The heavy lines of the linear DNA, bacterial chromosome, and cosmid are double-stranded DNA.)

Charon phages are selected simply by their ability to infect *E. coli* cells. As we mentioned, after manipulation, only λ DNA with a foreign insert is packaged because of the size requirement. Plasmids that contain foreign DNA can be selected through screening for antibiotic resistance. For example, a widely used cloning plasmid is named pBR322. (Plasmids are often named with the initials of their developers. The vector pBR322 was first described in a paper published in 1977 by authors F. Bolivar and R. Rodriguez, hence pBR.) Plasmid pBR322 contains genes for tetracycline and ampicillin resistance and various restriction sites. There is, for example, a *Bam*HI site in the tetracycline-resistance gene (fig. 13.8). After the ligating procedure, plasmids with and without foreign DNA will be present. *E. coli* cells are then exposed to this DNA mixture in the presence of calcium chloride; after taking up the DNA, the *E. coli* cells are plated on a medium without antibiotics. Replica-plating is done onto plates with one or both antibiotics. Colonies resistant to both antibiotics are composed of cells with plasmids having no inserts; those resistant only to ampicillin have a plasmid with an insert. Colonies resistant to neither antibiotic have cells with no plasmids.

### Blunt-End Ligation

Restriction endonuclease treatment may not suffice for cloning; an endonuclease may cut in the wrong place, say in the middle of a desired gene, or the foreign DNA may have been isolated by other methods, such as physical shearing. In these cases, several other methods of cloning can be used.

The most common method of joining foreign and vehicle molecules that do not have sticky ends is called **blunt-end ligation;** the phage enzyme, T4 DNA ligase, can join blunt-ended DNA. Blunt ends can be generated when segments of DNA to be cloned are created by physically breaking the DNA or by using certain restriction endonucleases, such as *Hin*dII (see fig. 13.2), that form blunt ends. Since the ligase is nonspecific about which blunt ends it joins, many different, unwanted products result from its action. Restriction enzymes that produce sticky ends are preferred for cloning.

A variation of blunt-end ligation uses **linkers**—short, artificially synthesized pieces of DNA containing a restriction endonuclease recognition site. When these linkers are attached to blunt pieces of DNA and then

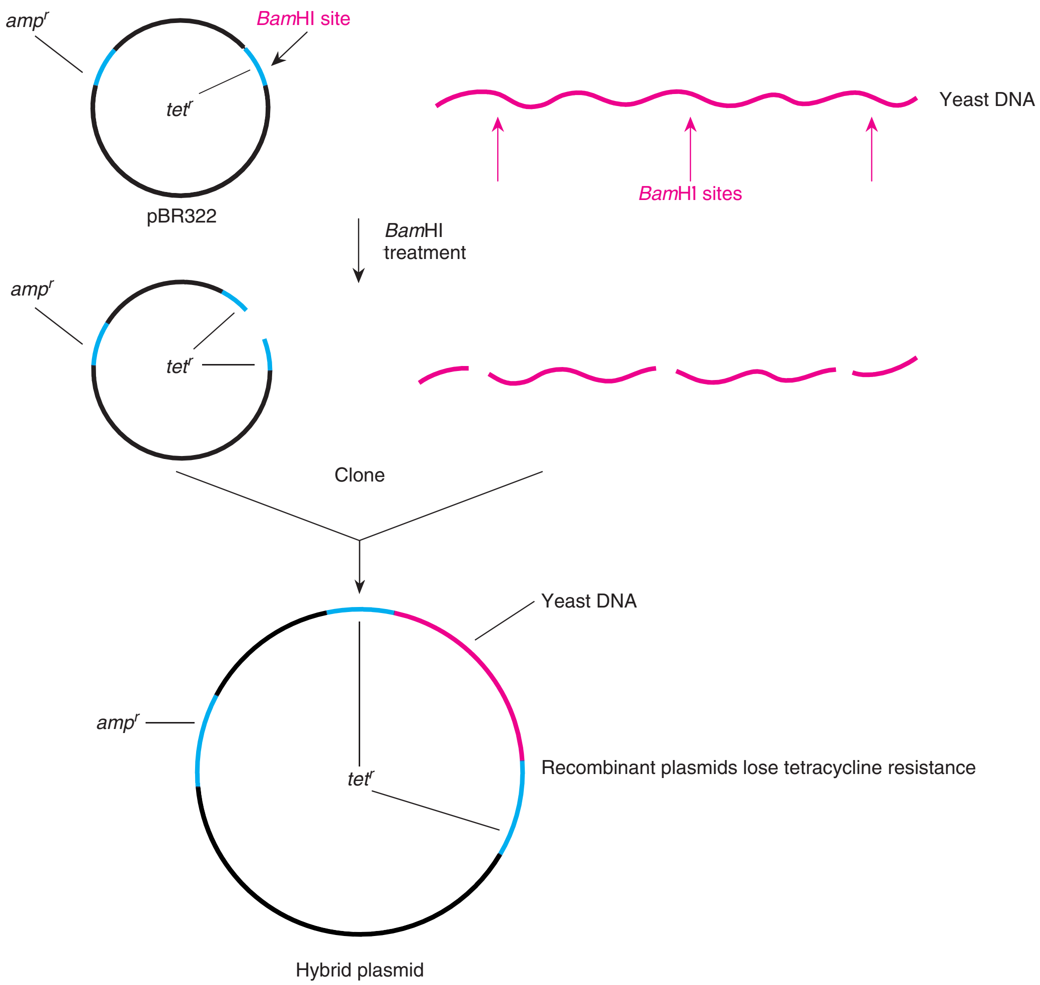


**Figure 13.8** *E. coli* plasmid pBR322. This plasmid carries two genes, $amp^r$ and $tet^r$, that confer resistance to ampicillin and tetracycline, respectively. A *Bam*HI restriction site occurs within the $tet^r$ gene. A cloned fragment within the $tet^r$ gene therefore destroys the tetracycline resistance. (Heavy *black, blue,* and *red* lines represent double-stranded DNA.)

treated with the appropriate restriction endonuclease, sticky ends are created. In figure 13.9, the linkers are twelve base-pair (bp) segments of DNA with an *Eco*RI site in the middle. They are attached to the DNA to be cloned with T4 DNA ligase. Subsequent treatment with *Eco*RI will result in DNA with *Eco*RI sticky ends.

DNA for cloning can be obtained generally in two ways: (1) a desired gene or DNA segment can be synthesized or isolated or (2) the genome of an organism can be broken into small pieces and the small pieces can be randomly cloned **(shotgun cloning).** Then the desired DNA segment must be "fished" out from among the various clones created. Let us look first at synthesizing or isolating a desirable gene before cloning it, and then look at the process of locating a desired gene after it has been cloned.

## Cloning a Particular Gene

### Creating DNA to Clone

To clone a particular gene (or DNA segment), a scientist must have a purified double-stranded piece of DNA containing that gene. There are numerous ways to obtain that DNA; several entail creating or isolating a single-stranded messenger RNA that is then enzymatically converted into double-stranded DNA. The problem is then reduced to obtaining the desired messenger RNA.

The messenger RNA for a particular gene can be obtained in several different ways, depending on the particular gene. If large quantities of the RNA from a particular cell are available, the RNA can be isolated directly. For example, mammalian erythrocytes have abundant quanti-

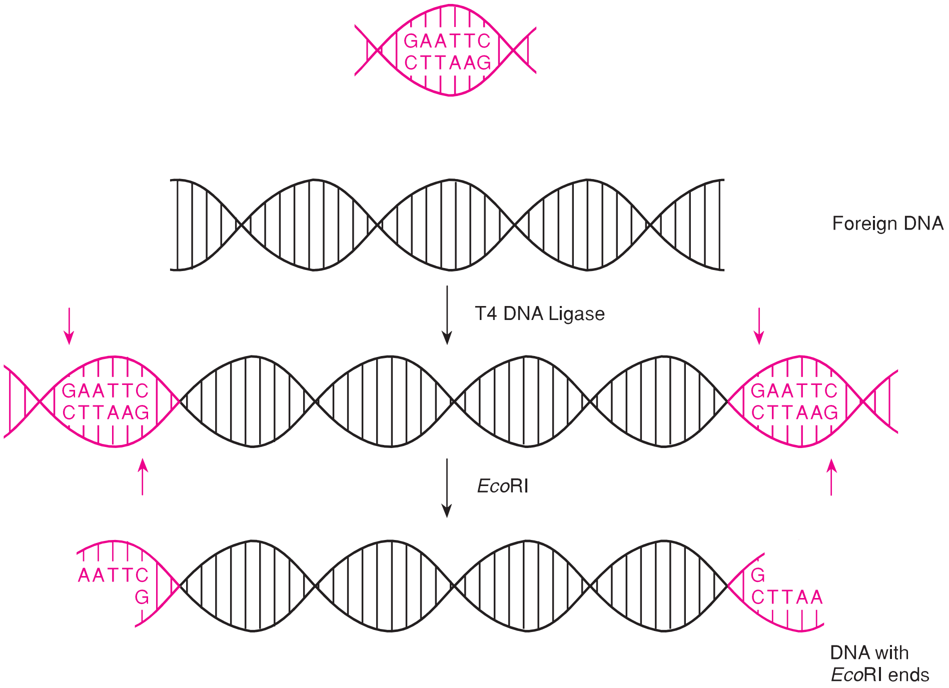


**Figure 13.9** Linkers: small segments of DNA with an internal restriction site. Linkers can be added to blunt-ended DNA by T4 DNA ligase. The restriction enzymes create DNA with ends that are compatible with any DNA cut by the same restriction enzyme (in this case, *Eco*RI).

ties of α- and β-globin messenger RNAs. Also, ribosomal RNA and many transfer RNAs are relatively easy to isolate in quantities adequate for cloning.

Double-stranded DNA for cloning is made from the purified RNA with the aid of the enzyme reverse transcriptase, isolated from RNA tumor viruses (see chapter 10). We describe here the conversion of RNA to DNA using a eukaryotic messenger RNA with a 3′ poly-A tail (fig. 13.10). In the first step, a poly-T primer is added, which base-pairs with the poly-A tail of the messenger RNA. This short, double-stranded region is now a primer for polymerase activity—a free 3′-OH exists. The primed RNA is then treated with the enzyme reverse transcriptase, which will polymerize DNA nucleotides using the RNA as a template. The result is a DNA-RNA hybrid molecule (fig. 13.10*c*).

The hybrid is now treated with the enzyme RNaseH, which creates random nicks in the RNA part of the RNA-DNA hybrid. These nicks provide the primer configuration for repair synthesis, the same repair done on Okazaki fragments when RNA primer is removed and replaced by DNA. Thus, the hybrid is treated with DNA polymerase I, which replaces each small RNA segment with DNA, base by base. Finally, the short DNA segments of the second DNA strand are united with DNA ligase (fig. 13.10*f*). The resulting double-stranded DNA is referred to as **complementary DNA (cDNA).** Hence, starting with a piece of single-stranded messenger RNA, we have generated a piece of double-stranded DNA. This piece can now be cloned using the blunt-end methods we have described.

If the RNA is not available in large enough quantities, it is possible to synthesize DNA in vitro if the amino acid sequence of its expressed protein is known. A possible nucleotide sequence can be obtained from the genetic code dictionary (see table 11.4) if the sequence of amino acids is known from the protein product of the gene. This method will probably not re-create the original DNA because of the redundancy in the genetic code. In other words, any one of six different codons could have coded a particular leucine in a protein. Despite an element of guesswork, it is possible to synthesize a piece of DNA that will code for a particular protein. Currently, automated machines that add one base at a time in ten-minute cycles can synthesize DNA sequences of over one hundred bases.

You will notice that the methods we have described, making cDNA or synthetic DNA using the genetic code dictionary, produce DNA missing the gene's promoter and its transcriptional control sequences as well as other untranslated areas of the DNA (introns). If it is desirable to clone an intact gene with its promoter and introns, then cloning can be done by creating random pieces of

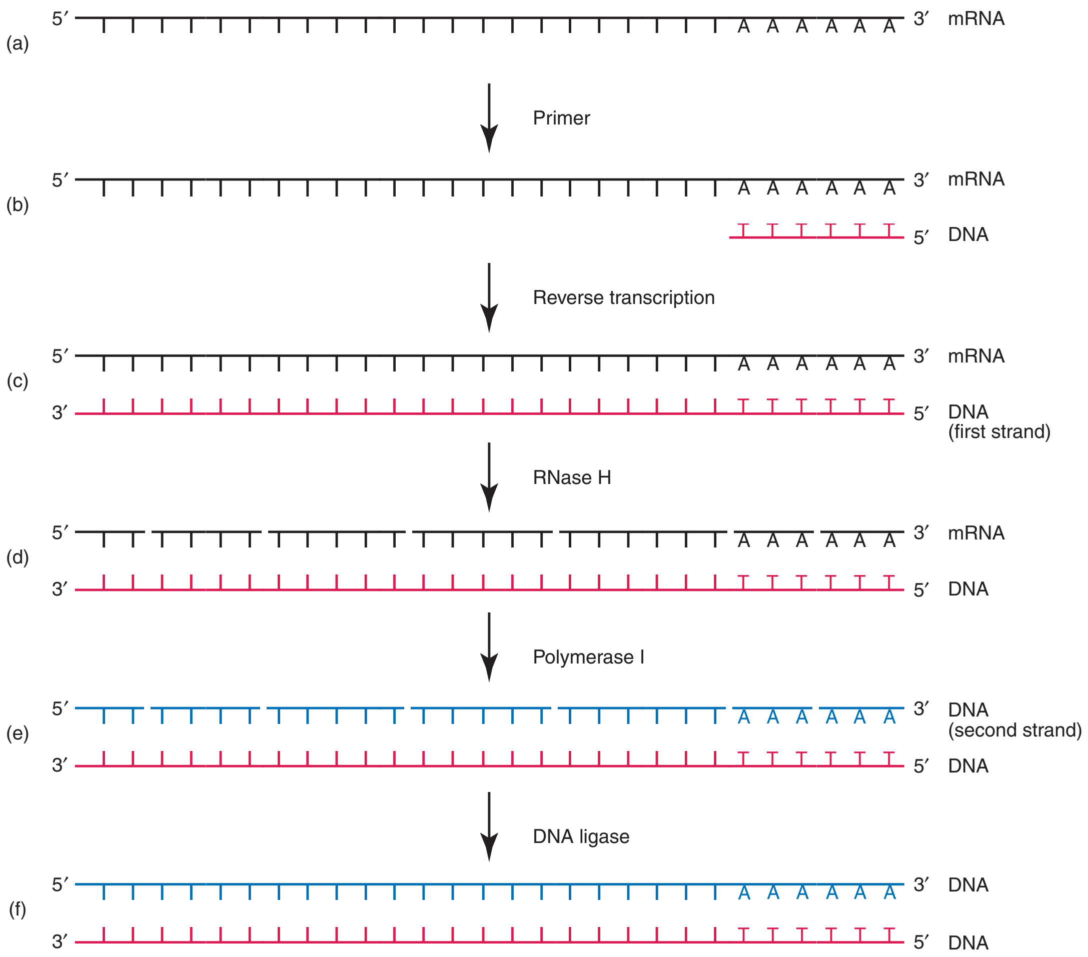


**Figure 13.10** (*a*) A messenger RNA, shown in *black,* begins as a single strand. (*b*) A poly-T DNA segment (*red*) is added as primer; it complements the 3′ poly-A tail of the eukaryotic messenger RNA. (*c*) Reverse transcriptase acts on this primed configuration to synthesize a single strand of DNA from the RNA template. (*d*) The RNA is then nicked randomly by RNase H. (*e*) The RNA segments are then replaced by DNA (*blue*) by the action of DNA polymerase I. (*f*) After DNA ligase treatment, the final result is double-stranded complementary DNA (cDNA).

the genome. The gene of interest can be found either before or after cloning it, although it is usually done after cloning.

### Creating a Genomic Library

When cDNA or synthetic DNA cannot be used for cloning, the total DNA of an organism can be broken into small pieces to isolate the desired gene or DNA fragment. The desired DNA can be isolated either before or after cloning. This DNA is referred to as genomic DNA to differentiate it from cDNA.

If the original DNA is isolated before cloning, then only that DNA need be cloned. Alternatively, a "shotgun" approach can be used to clone a sample of the entire genome of an organism (in small pieces, of course), creating a **genomic library,** a set of cloned fragments of the original genome of a species (fig. 13.11). In a genomic library, a desired gene can be located after it is already cloned.

## Southern Blotting

When DNA segments are generated randomly, usually by endonuclease digestion, a desired gene must be located. As mentioned, we can look for the gene either before or after it is cloned. We consider first the procedure for locating a specific gene in a DNA digest, before the DNA has been cloned.

To locate a specific gene in the midst of a DNA digest, one must have a specific **probe.** Probes are generally nucleic acids with sequences that precisely locate a complementary DNA sequence by hybridization. The probes are labeled so they can be identified later with autoradiography or **chemiluminescent techniques** (techniques in which tags are used that fluoresce under ultraviolet or laser light). Thus, if we wish to locate the gene for β-globin, we could use radioactively labeled β-globin messenger RNA or radioactively labeled cDNA. RNA-DNA or DNA-DNA hybrids would form between the specific gene and the radioactive probe. Autoradiography or chemiluminescence would then locate the radioactive probe.

Let us assume that we wanted to clone the rabbit β-globin gene. First, we would create a restriction digest of rabbit DNA (fig. 13.12). We would then subject this digest to electrophoresis on agarose to separate the various fragments according to size. Agarose is a good medium for separating DNA fragments of a wide variety of sizes.

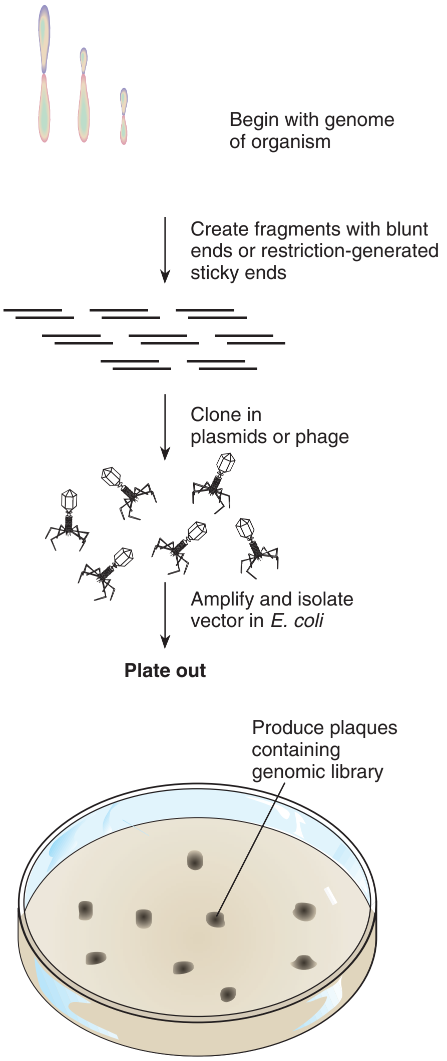

Begin with genome of organism
Create fragments with blunt ends or restriction-generated sticky ends
Clone in plasmids or phage
Amplify and isolate vector in *E. coli*
Plate out
Produce plaques containing genomic library

**Figure 13.11** Creating a genomic library using the shotgun approach in creating inserts. First, the genome is fragmented. The fragments are then cloned randomly in vectors. The collection of these vectors is referred to as a genomic library.

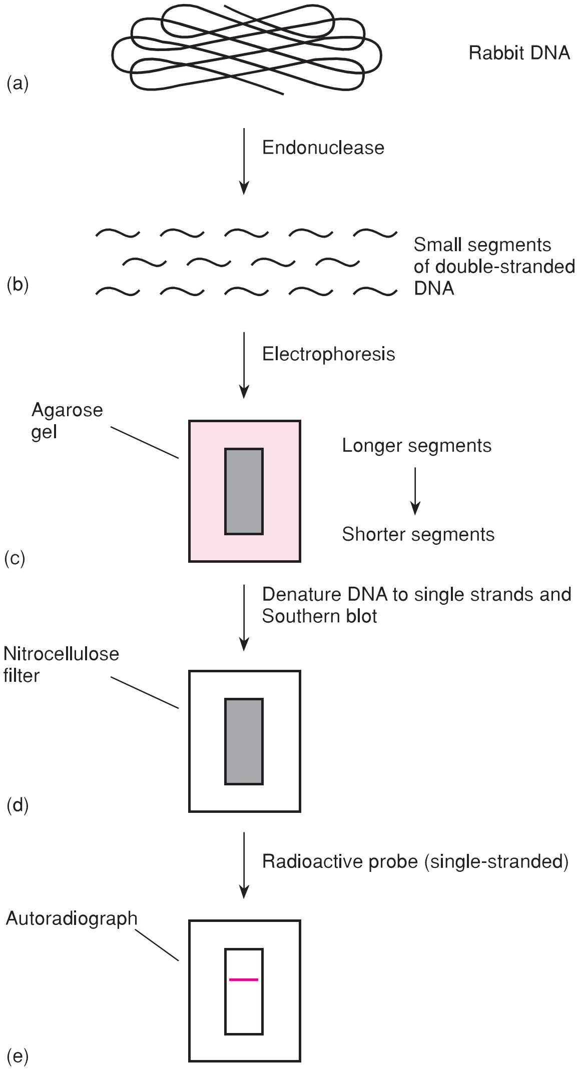

(a)
Rabbit DNA
Endonuclease
(b)
Small segments of double-stranded DNA
Electrophoresis
Agarose gel
Longer segments
Shorter segments
(c)
Denature DNA to single strands and Southern blot
Nitrocellulose filter
(d)
Radioactive probe (single-stranded)
Autoradiograph
(e)

**Figure 13.12** Locating the rabbit β-globin gene within a DNA digest using the Southern blotting technique. The rabbit DNA (*a*) is segmented with a restriction endonuclease (*b*) and then electrophoresed on agarose gels (*c*). Southern blotting transfers the DNA to nitrocellulose filters (*d*). Finally, a radioactive probe (β-globin messenger RNA) locates the DNA fragment with the β-globin gene after autoradiography (*e*).

Edwin M. Southern (1938– ).
(Courtesy of Edwin Southern.)

In a digest of this kind, however, there are usually so many fragments that the result is simply a smear of oligonucleotides, from very small to very large. To proceed further, we have to transfer the electrophoresed fragments to another medium for probing, or the DNA fragments would diffuse out of the agarose gel. Nitrocellulose filters or nylon membranes are excellent for hybridization because the DNA fragments bind to these membranes and will not diffuse out. The transfer procedure, first devised by E. M. Southern, is called **Southern blotting.** In this technique, the double-stranded DNA on the agarose gel is first denatured to single-stranded DNA, usually with NaOH. Then the agarose gel is placed directly against a piece of nitrocellulose filter, and the resulting sandwich is placed agarose-side-down on a wet sponge. Dry filter paper placed against the nitrocellulose side wicks fluid from the sponge, through the gel, and past the nitrocellulose filter, carrying the DNA segments from the agarose to the nitrocellulose (fig. 13.13). NaOH is used as the transfer solution in the tray. The DNA digest fragments are then permanently bound to the nitrocellulose filter by heating. DNA-DNA hybridization takes place on the filter. (A similar technique can be performed on RNA, which is called, tongue-in-cheek, **northern blotting.** Immunological techniques, not involving nucleotide complementarity, can be used to probe for proteins in an analogous technique called **western blotting.**)

A labeled probe can be obtained in several different ways. In this example, the easiest way to obtain a radioactive probe would be to isolate β-globin messenger RNA from rabbit reticulocytes and construct cDNA using the reverse-transcriptase method described. The deoxyribonucleotides used during reverse transcription are then synthesized to contain radioactive phosphorus, $^{32}$P. As figure 13.12 shows, after hybridization, a single radioactive band locates a DNA segment with the β-globin gene. Note that the probe, originating from messenger RNA, will lack the introns present in the gene. However, probing is successful as long as there are complementary regions in the two nucleotide strands.

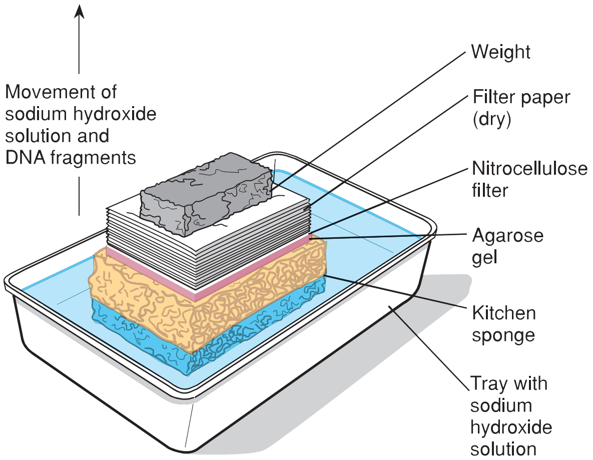

Movement of sodium hydroxide solution and DNA fragments
Weight
Filter paper (dry)
Nitrocellulose filter
Agarose gel
Kitchen sponge
Tray with sodium hydroxide solution

**Figure 13.13** Arrangement of gel and filters in the Southern blotting technique. The NaOH buffer is drawn upwards by the dry filter paper, transferring the DNA from the agarose gel to the nitrocellulose filter.

To clone the β-globin gene, a second agarose gel would be run with a sample of the digest used in figure 13.12. That gel, not subject to DNA-DNA hybridization, would have the β-globin segment in the same place. The band, whose location is known from the autoradiograph, could be cut out of the agarose gel to isolate the DNA. We could then clone the DNA by methods discussed earlier in the chapter.

## *Probing for a Cloned Gene*

### *Dot Blotting*

The methods we have described are also useful in locating genes already cloned within plasmids, for example, after a genomic library has been constructed. In this case, electrophoresis and Southern blotting are not needed since we will be probing for a particular sequence of DNA already cloned rather than DNA segments within a digest.

For example, the DNA of a human-mouse hybrid cell line was cloned in order to locate human DNA. In this case, a hybrid cell line had only one human chromosome, chromosome 20. In order to locate DNA from that chromosome, probes were used that were isolated from human chromosomes. The probes were radioactively labeled. Meanwhile, 288 *E. coli* colonies, each containing a hybrid plasmid, were grown and transferred directly to a nitrocellulose filter. In preparation for probing, the cells were lysed and their DNA denatured. The plasmid DNA within the cells of each clone was then hybridized with the radioactive probes. Figure 13.14 is an autoradiograph of the 288 clones. The two dark spots indicate clones carrying DNA from human chromosome 20 (those clones "light up" autoradiographically). This technique, hybridization of cloned DNA without an electrophoretic-separation step, is referred to as **dot blotting.**

These techniques can also be carried out without the grid arrangement of colonies by using replica plating as described in chapter 7. Thus, a specific gene can be located after shotgun cloning.

### Western Blotting

An entirely different method used to locate particular cloned genes utilizes the actual expression of the cloned genes in the plasmid-containing cells. If a eukaryotic gene is cloned in an *E. coli* plasmid downstream from an active promoter, that gene may be expressed (transcribed and translated into protein). Plasmids that allow the expression of their foreign DNA are termed **expression vectors.** There are many problems with this technique because bacteria normally would not express eukaryotic genes. However, special vectors have been developed in which the cloning site is just downstream from a promoter. The eukaryotic gene thus becomes part of the prokaryotic gene, producing a fusion protein, usually with only a few amino acids from the prokaryote. Of course, the eukaryotic gene must be in the correct orientation and in the correct codon reading frame for appropriate translation, meaning that the success rate of this technique is relatively low.

A particular protein product can be located by western blotting, a method completely analogous to either Southern blotting or dot blotting. In this technique, probing is done with antibodies specific for a particular protein, rather than using a radioactive oligonucleotide probe. A second antibody, specific to the first and labeled with a marker, usually fluorescent, locates the first antibody. For example, assume we are looking for the expression of a particular protein in the clones of figure 13.14. The clones would be transferred to a nitrocellulose membrane, where they would be lysed (e.g., with chloroform vapor). Then an antibody, specific for the particular protein, would be applied. A second antibody, specific for the first antibody and labeled with a fluorescent marker, would be applied to the filters. Fluorescence of the second antibody would locate the presence of the first antibodies and thus indicate which of the clones is expressing the particular gene (fig. 13.15).

## Eukaryotic Vectors

The work we have described so far involves introducing chimeric plasmids into bacteria, primarily *E. coli.* However, there are several reasons why we want to extend these techniques to eukaryotic cells (box 13.1). First, a prokaryote like *E. coli* is not capable of fully expressing some eukaryotic genes since it lacks the enzyme systems necessary for some posttranscriptional and posttranslational modifications such as intron removal and some protein modification. Second, we also wish to study the organization and expression of the eukaryotic genome in

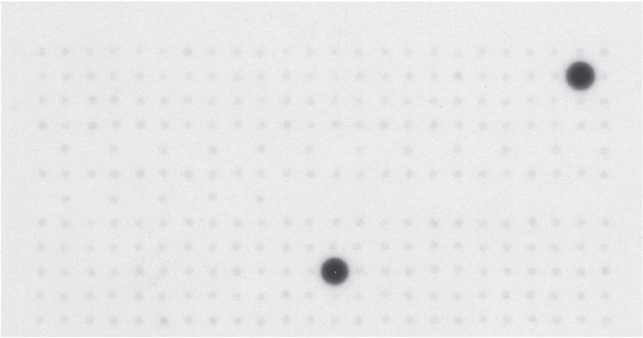


**Figure 13.14** Dot blot autoradiograph of 288 clones of DNA from a mouse-human hybrid cell line. After lysing samples of each clone on a nitrocellulose filter, the investigator hybridized the clones with radioactive probes for human-specific sequences. The two *dark spots* indicate clones carrying human DNA. The slight background radiation in most other spots provides the spot pattern needed to orient the investigator. (Source: Courtesy of Nick O. Bukanov.)

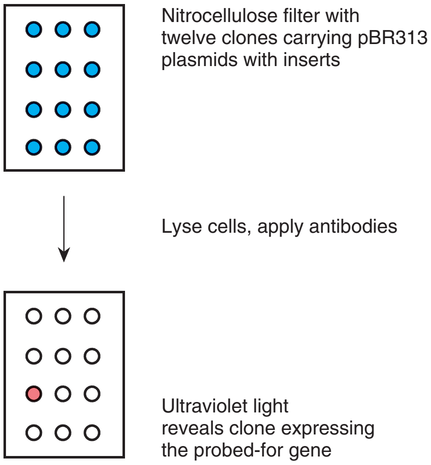
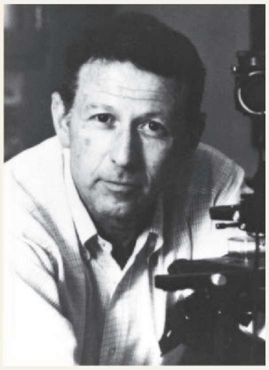

Nitrocellulose filter with twelve clones carrying pBR313 plasmids with inserts
Lyse cells, apply antibodies
Ultraviolet light reveals clone expressing the probed-for gene

**Figure 13.15** Western blot technique is used to locate an expressed protein from among many clones. Clones that may carry the expressed protein are lysed. Tagged antibodies are applied to locate the protein; a second antibody locates the first antibody either through a fluorescent color marker, as shown here, or with autoradiography.

## BOX 13.1



Paul Berg (1926– ).
(Courtesy of Dr. Paul Berg.)

### Ethics and Genetics

### *The Recombinant DNA Dispute*

Paul Berg shared the 1980 Nobel Prize in chemistry for creating the first cloned DNA molecule, a hybrid λ phage that contained the genome of the simian tumor virus, SV40. The fact that he could do this work was worrisome to many people, himself included. The recombinant DNA dispute was underway. Berg voluntarily stopped inserting tumor virus genes into phages that attack the common intestinal virus *E. coli.*

People continue to worry about the dangers of working with recombinant DNA. One immediate and obvious concern is that cancer or toxin genes will "escape" from the laboratory. In other words, recombinant DNA technology could create a bacterium or plasmid that contained toxin or tumor genes. The modified bacterium or plasmid could then accidentally infect people. A 1974 report by the National Academy of Sciences led to a February 1975 meeting, which took place at the Asilomar Conference Center south of San Francisco. Berg convened this meeting, which over one hundred molecular biologists attended. The recommendations of the Asilomar Committee later formed the basis for official guidelines developed by the National Institutes of Health (NIH). In essence, NIH established guidelines of containment.

Containment means erecting physical and biological barriers to the escape of dangerous organisms. The NIH guidelines defined four levels of risk, from minimal to high, and four levels of physical containment for them (called P1 through P4). The most hazardous experiments, dealing with the manipulation of tumor viruses and toxin genes, require extreme care, which included negative-pressure air locks to the laboratory and experiments done in laminar-flow hoods, with filtered or incinerated exhaust air.

Biological containment means developing host cells and manipulated vectors that are incapable of successful reproduction outside the lab, even if they escape. High-risk work was done with host cells or vectors that were modified. For example, a bacterium of the *E. coli* strain EK2 cannot survive in the human gut because it has mutations that do not permit it to synthesize thymine or diaminopimelate. The lack of thymine-

---

vivo (in the living system), something we can only accomplish by working directly with eukaryotic cells. Finally, we wish to learn how to manipulate the genomes of eukaryotes for medical as well as economic reasons. To these ends, we discuss eukaryotic plasmids and the direct manipulation of eukaryotic genomes in vivo.

### *Yeast Vectors*

Yeasts, small eukaryotes that can be manipulated in the lab, like prokaryotes, have been studied extensively. Baker's yeast, *Saccharomyces cerevisiae,* has a naturally occurring plasmid. In addition, bacterial plasmids have been introduced into yeast. Unfortunately, the cells tend to lose these plasmids. This tendency has been overcome, however, by constructing bacterial plasmids that contain a yeast centromere (CEN) and the origin of yeast DNA replication (ARS for autonomously replicating sequence; fig. 13.16). The yeast then carries the plasmids from one generation to the next. The plasmids can have telomeric sequences inserted, and they can then be made linear by cutting the telomeric sequences with endonuclease. Alternatively, the plasmids can be linearized first, and then have telomeric sequences added to their ends. The plasmids are then called **yeast artificial chromosomes (YACs).** The particular advantage YACs have is that they are capable of accepting very large pieces of inserted DNA. Remember that a cosmid can hold about 50 kb; a YAC can hold as much as 800 kb or more. The ability to clone this much DNA is valuable when working with large eukaryotic genes and in the Human Genome Project (see the section with this title later in the chapter).

Recombinant DNA studies in yeast have increased our knowledge about gene regulation in eukaryotes, about how the centromere works, and about the way in which the tips of eukaryotic linear chromosomes are replicated. In addition, YACs have allowed us to analyze and sequence very large segments of eukaryotic DNA.

synthesizing ability is lethal because the cell cannot replicate its DNA. The diaminopimelate is a cell-wall constituent; without it, the cells burst. These bacteria also carry mutations that make them extremely sensitive to bile salts. Thus, if by accident the cells were to escape, they would pose virtually no threat. The plasmids used for recombinant research were modified so that they could not be transferred from one cell to the next. Again, if containment failed, neither the host cells nor their plasmids would survive.

In 1979, the guidelines were relaxed. Although it was wise to be cautious, it appears that initial fears were unwarranted. Recombinant DNA work now seems to pose little danger: Containment works very well, and engineered bacteria do very poorly under natural conditions. *E. coli* has been living in mammalian guts for millions of years, so it has had numerous opportunities to incorporate mammalian DNA into its genome (intestinal cells are dying and sloughing off into the gut all the time). No "Andromeda strain" has arisen, nor do we foresee one in the future.

Current concern is focused on the acceptability of genetically modified crops (GM crops). As we will discuss later, one fourth of American cropland is planted with genetically modified crops, modified mainly for insect resistance. These modifications have curtailed our use of insecticides. (For cotton and corn, for example, liquid insecticide use dropped by 3.6 million liters and powdered insecticide by 300,000 kilograms in 1999.) However, people are concerned with the effects these modifications might have on natural ecosystems: How many valuable insects will be killed by mistake? Although Third World countries are desperate for these technologies, the United States, European and Asian trading partners are demanding that the crops we export be genetically unmodified. Farmers are also concerned that genetically modified crops have been modified to be sterile (so called "terminator technology") so that farmers would need to buy new seeds each year.

More recently, the recombinant DNA dispute has taken a whole new twist. It now has surfaced as a conflict between academic freedom and industrial secrecy. It seems that recombinant DNA technology is very lucrative. Numerous academic scientists have either begun genetic engineering companies or become affiliated with pharmaceutical companies. However, the philosophies of private enterprise and academia are often in conflict. Academic endeavors are presumably open, with free exchange of information among colleagues, whereas private enterprise entails some degree of secrecy, at least until patents are obtained to protect the investments of the companies. Thus, a basic conflict can arise for scientists trained in gene cloning. The conflict has been prevalent since late 1980, when the first patent for recombinant DNA techniques was awarded to Stanford University and the University of California. When, in April 2000, United States President Bill Clinton and British Prime Minister Tony Blair issued a joint statement asking that human genome data not be patented, the American stock market took a major downturn. This is a tumultuous time for biotechnology.

### Animal Vectors

The vehicle most commonly used in higher animals is the DNA tumor virus SV40. (SV, or simian vacuolating virus, was first isolated in monkeys; however, it can transform normal mouse, rabbit, and hamster cells. Unlike the use of the word *transformation* in bacteria, transformation in eukaryotes refers to the changing of a normal cell into a rapidly growing, cancerous one.) SV40 is an icosahedral particle with a small (5,224 base pairs) chromosome, which is a circular, double-stranded DNA molecule.

Like λ vectors, SV40 virions allow foreign DNA to replace part of their DNA. The viruses can then be used in recombinant DNA studies in one of two ways (fig. 13.17). They can replicate and complete their life cycle with the help of nonrecombinant viruses, or they can replicate in the host without making active virus particles by existing as circular plasmids in the cytoplasm or by integrating into the host's chromosomes. SV40 has become a valuable tool in mammalian genomic studies. For example, the rabbit β-globin gene was cloned in SV40, and enhancer sequences (see chapter 10) were discovered in SV40. DNA tumor viruses have also deepened our understanding of transformation in eukaryotes (oncogenesis).

### Plant Vectors

The best-studied system for introducing foreign genes into plants is the naturally occurring crown gall tumor system. The soil bacterium *Agrobacterium tumefaciens* causes tumors, known as crown galls, in many dicotyledonous plants (fig. 13.18). In essence, the crown gall is made of transformed plant cells. These cells have been transformed by a plasmid within the bacterium called the *tumor-inducing,* or *Ti,* plasmid. Transformation occurs when a piece of the plasmid called T-DNA (for transferred DNA) is integrated into the chromosome of the plant host. Crown gall cells produce amino acid derivatives, termed *opines,* that the *A. tumefaciens* cells use. By manipulating this

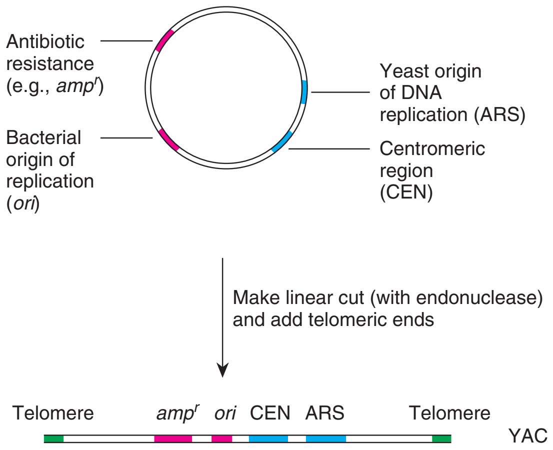

Antibiotic resistance (e.g., *amp*$^r$)
Bacterial origin of replication (*ori*)
Yeast origin of DNA replication (ARS)
Centromeric region (CEN)
Make linear cut (with endonuclease) and add telomeric ends
Telomere
*amp*$^r$
*ori*
CEN
ARS
Telomere
YAC

**Figure 13.16** *Escherichia coli* plasmid pBR322 modified for use in yeast. This plasmid survives and replicates in both yeast and *E. coli* because it contains the origin of replication for both, as well as a yeast centromeric region (CEN). When it is made linear and telomeres are added, the yeast artificial chromosome (YAC) becomes suitable for cloning large pieces of DNA.

system, geneticists have begun to understand the transformation process in plants as well as to develop a manipulatable system for introducing foreign genes into plants.

The study of genetics in plants has been boosted a great deal by the availability of model organisms similar to *E. coli,* yeast, and fruit flies. Recently, much attention has focused on the meadow weed, *Arabidopsis thaliana* (fig. 13.19). This small plant is ideal for studying plant genetics because its genome is small, approximately 100 million base pairs located in only five chromosomes ($2n = 10$). This is only about five times the genome of yeast or twenty times the genome of *E. coli.* Thus, in terms of genome size, it is quite manageable. *A. thaliana* has joined the ranks of organisms whose genomes have been sequenced. The plants are easy to grow in very large numbers, and each plant produces as many as ten thousand seeds. Hence, this organism compares very favorably with fruit flies and yeast for studying questions of gene control in a eukaryote, in this case a plant.

## Expression of Foreign DNA in Eukaryotic Cells

Foreign DNA can be introduced into eukaryotic cells in methods similar to bacterial transformation. However, the process in eukaryotes is called **transfection** because, as we described, the term *transformation* in eukaryotes is used to mean cancerous growth. Eukaryotic organisms that take up foreign DNA are referred to as **transgenic.** Most of the techniques described here transcend taxonomic lines.

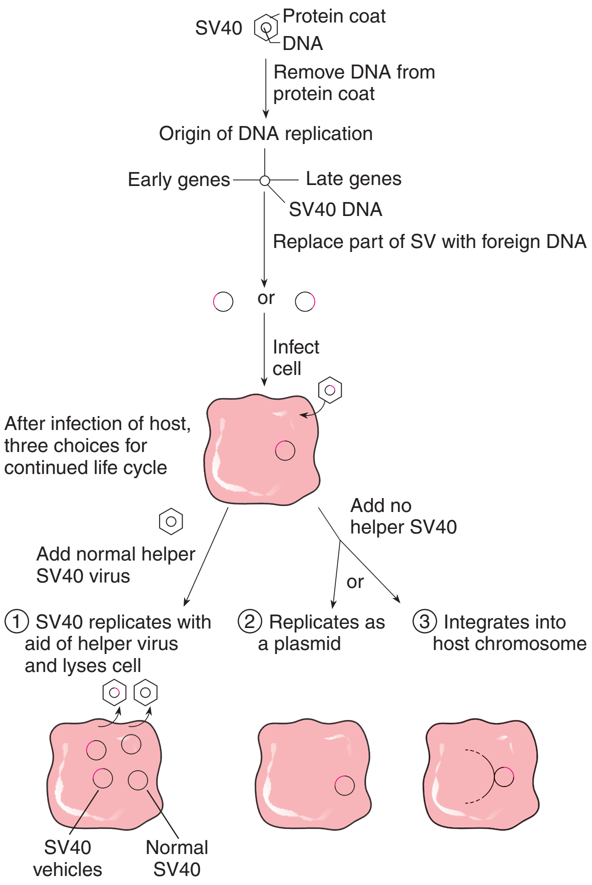

SV40
Protein coat
DNA
Remove DNA from protein coat
Origin of DNA replication
Early genes
Late genes
SV40 DNA
Replace part of SV with foreign DNA
or
Infect cell
After infection of host, three choices for continued life cycle
Add normal helper SV40 virus
Add no helper SV40
or
① SV40 replicates with aid of helper virus and lyses cell
② Replicates as a plasmid
③ Integrates into host chromosome
SV40 vehicles
Normal SV40

**Figure 13.17** SV40 virus can be used as a gene cloning vehicle. Although part of the virus is replaced by inserted DNA during cloning, it can still replicate with the aid of normal helper viruses (nonrecombinant SV40). Without the aid of helper viruses, it can either replicate as a plasmid or integrate into the host chromosome.

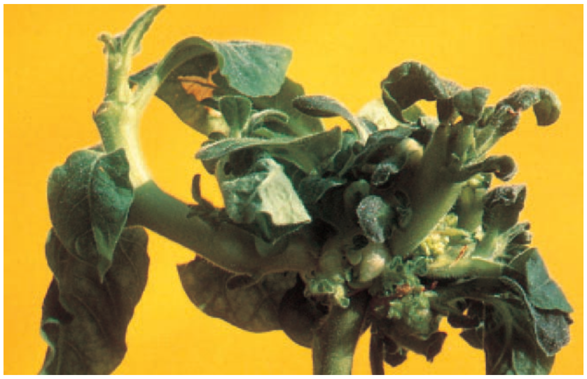


**Figure 13.18** Crown gall on tobacco plant (*Nicotiana tabacum*) produced by *Agrobacterium tumefaciens* containing Ti plasmids. (Courtesy of Robert Turgeon and B. Gillian Turgeon, Cornell University.)

Animal cells, or plant cells with their walls removed (protoplasts), can take up foreign chromosomes or DNA directly from the environment with a very low efficiency (in the presence of calcium phosphate). Directly injecting the DNA greatly improves the efficiency. For example, transgenic mice are now routinely prepared by injecting DNA into either oocytes or one- or two-celled embryos obtained from female mice after appropriate hormonal treatment (fig. 13.20). After injection of about 2 picoliters ($2 \times 10^{-12}$ liters) of cloned DNA, the cells are reimplanted into the uteruses of receptive female hosts. In about 15% of these injections, the foreign DNA incorporates into the embryo. Transgenic animals are used to study the expression and control of foreign eukaryotic genes. In 1988, a transgenic mouse prone to cancer was the first genetically engineered animal to be patented. This mouse provides an excellent model for studying cancer (see chapter 16). (A controversy arose as to whether engineered higher organisms should be patentable; currently they are.) Mice have already been successfully transfected with a rat growth-hormone gene (fig. 13.21), and transgenic sheep have been produced that express the gene for a human clotting factor. The latest recombinant DNA dispute arises from the cloning of sheep in 1997 (box 13.2).

Transfection can also be mediated by retroviruses (RNA viruses containing the gene for reverse transcriptase). For example, a retroviral vector infected and repaired human white blood cells lacking the enzyme

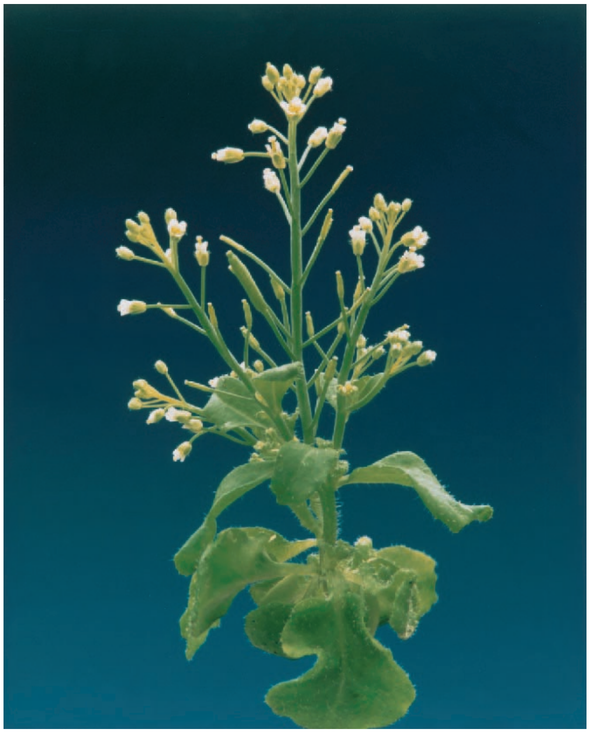


**Figure 13.19** A dwarf form of the plant *Arabidopsis thaliana.* (Source: *Science,* Vol. 243, March 10, 1989, cover. ©1989 AAAS, Washington, D.C. Photo by DeVere Patton. Courtesy of E. I. DuPont de Nemours and Company.)

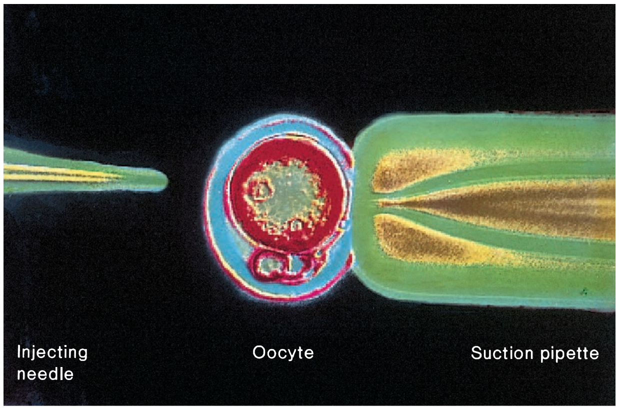


**Figure 13.20** Injection of DNA into the nucleus (germinal vesicle) of a mouse oocyte. The oocyte is held by suction from a pipette. (© John Gardon/Phototake.)

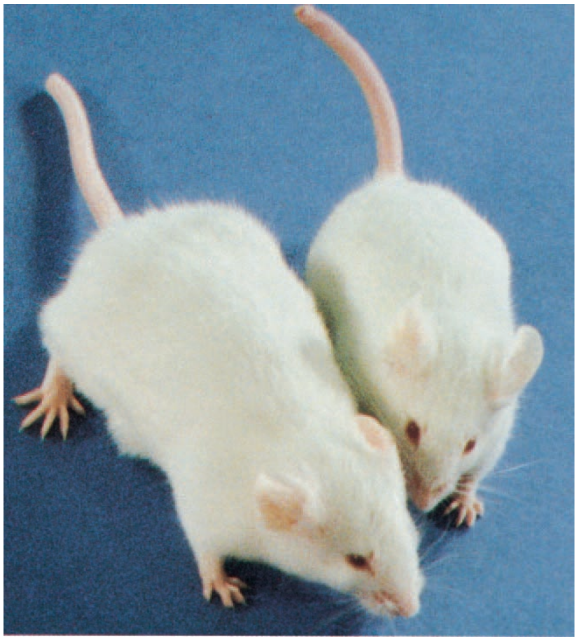


**Figure 13.21** Mouse littermates. The larger one is a transgenic mouse containing the rat growth-hormone gene. (Source: Richard D. Palmiter, Ralph L. Brinster, et al., "Dramatic growth of mice that develop from eggs microinjected with metallothionein growth hormone fusion genes," *Nature* 300, 16 December 1982, cover. Photograph by Dr. Ralph L. Brinster. Copyright © 1982, Macmillan Magazines Ltd.)

## BOX 13.2

## Ethics and Genetics

### *Cloning Dolly*

Fiction writers in the past have created stories in which scientists cloned one person to create numerous copies. The themes of these stories have varied from the cloning of Adolph Hitler to the cloning of a very busy man to help him fulfill his day-to-day obligations. The possibility of those scenarios came a bit closer to reality in February 1997 when a group of scientists from the Roslin Institute and PPL Therapeutics, both in Edinburgh, Scotland, reported in *Nature* magazine that they had successfully cloned a sheep from a cell taken from the udder of a six-year-old ewe. The cloned lamb was named Dolly (fig. 1). In the past, genetically identical animal embryos had been created only with amphibian cells, and those created from adult nuclei had never successfully reached adulthood. Cloning in which the nuclei came from fetal cells or cells from cell lines had been successful before in mammals.

To clone an animal, it is necessary to begin with an egg, the only cell known to initiate and support development. In order to clone an individual, using the word *clone* to mean create a genetically identical copy, it is necessary to get an egg without a nucleus and then to transplant a nucleus of known origin. Techniques for nuclear transplantation had been worked out with frogs and toads in the 1950s. The Scottish scientists succeeded in obtaining sheep eggs, enucleating them (removing their nuclei), and then transferring in donor nuclei by fusing the donor cells and the enucleated eggs with an electrical pulse. The electrical pulse also initiated development of the egg. Although only one pregnancy of the twenty-nine initiated was successful, the lamb that was born seems normal in every way; it has since produced offspring.

Others had tried this type of experiment with many types of animals, including mice. They were not successful for numerous reasons. The most likely explanation for the recent success, according to the scientists, is that the donor cells were kept in a nongrowth phase for several days, which may have synchronized them with the oocyte. Thus, the nucleus and the oocyte were at the same stage of the cell cycle and thus compatible. Other reorganizations that had to take place in the donor chromosomes are not really known for certain, but one thing is clear: the nucleus of an adult cell in the sheep has all of the genetic material needed to support normal growth and development of an egg. (The work has since been repeated with goats, cattle, and mice.)

There are numerous ramifications to the success of this work. First, mammal cloning could become a routine procedure. This would allow us to study mammalian development and to replicate genetically identical individuals, particularly transgenic animals that would have particular genomes of value. We can also use these techniques to study aging, since an "old" nucleus is initiating the development of a new organism. Also of interest is the interaction of a particular genome with a particular cytoplasm, since the cytoplasm contains not only the materials needed for early development, but also cell organelles, including mitochondria that have their own genetic material.

Finally, ethical issues must be considered if this technique is successful with human beings. Parents might wish to clone a deceased child or to obtain an immunologically compati-

adenosine deaminase. A retrovirus responsible for a form of leukemia in rodents, the Moloney murine leukemia virus, was engineered so that all the viral genes were removed and replaced with an antibiotic marker (neomycin resistance) and the human adenosine deaminase gene. The virus binds to the cell surface and is taken into the cell, its RNA is converted to DNA by reverse transcription, and the DNA is incorporated into one of the cell's chromosomes. It is not possible for this highly modified virus to attack and damage the cells unless a helper virus is added. Unlike the SV40 viruses in figure 13.17, the modified Moloney viruses cannot initiate a successful infection without the helpers because vital genes have been removed.

Three other recent techniques deliver recombinant DNA to eukaryotic cells: electroporation, liposome-mediated transfer, and "biolistic" transfer. In **electroporation,** exogenous DNA is taken up by cells subjected to a brief exposure of high-voltage electricity. Presumably, this electric field creates transient micropores in the cell membrane, allowing exogenous DNA to enter.

Liposome-mediated transfection is a technique that encapsulates foreign DNA in artificial membrane-bound vesicles called **liposomes.** The liposomes are then used to deliver their DNA to target cells. In one experiment, 50% of mice injected with these DNA-containing liposomes were successfully transfected—they expressed the proteins the transfecting DNA encoded.

Last are techniques developed to deliver foreign DNA into mitochondria and chloroplasts. These have proven difficult targets for genetic engineering because, ble sibling for a child who needs an organ or bone marrow transplant. Others might oppose cloning based on their religious and moral convictions. In response to these latter considerations, President Bill Clinton urged Congress to ban the cloning of human beings in the United States in 1997 for at least five years.

**Figure 1** Dolly is the first cloned sheep produced by the transfer of a nucleus from the cell of an adult sheep. (AP/Wide World Photos.)

among other reasons, they have double-membrane walls that have not proven amenable to delivery of recombinant DNA. Recently, transfection has been successful in both mitochondria and chloroplasts using a **biolistic** (biological ballistic) process, literally shooting recombinant DNA coated on tungsten microprojectiles into these organelles.

### *Knockout Mice*

Normally, a gene used to transfect mice is incorporated randomly in the mouse genome. However, in about one in one thousand experiments, the gene replaces the normal gene by a process similar to meiotic recombination (*homologous recombination;* see chapter 12). With this process in mind, geneticists have been able to select for homologous recombination; by transfecting with defective genes, they have created mice without working copies of a particular gene. The mice produced are called **knockout mice,** and they give geneticists the opportunity to study the phenotype of an animal that lacks a particular gene.

The geneticist first creates a vector with the modified gene in question. In addition, flanking regions to that gene are added so that homologous recombination can occur. Finally, two antibiotic genes are introduced so that selection for successful transfection takes place. Within the flanking regions, the gene for neomycin resistance (*neo*$^r$) is inserted; its product inactivates the antibiotic neomycin (fig. 13. 22). Outside of the flanking regions, the gene for thymidine kinase *(tk)* is inserted. This gene phosphorylates the drug *gancyclovir;* the phosphorylated gancyclovir is a nucleotide analogue that is incorporated during DNA synthesis, killing the cell. Thus, the combination of the *tk* gene and gancyclovir is lethal; without the *tk* gene, gancyclovir is harmless. If cells are exposed to both drugs, neomycin and gancyclovir, normal cells will be killed by neomycin, cells with the *tk* gene will be killed by gancyclovir, and only the cells with the *neo$^r$* gene but lacking the *tk* gene will survive. These alternative outcomes allow us to select the cells in which homologous recombination took place (fig. 13.22).

Cells that did not incorporate the vector will die from the effects of neomycin (they are neomycin sensitive). Cells that randomly took up the vector DNA by nonhomologous recombination will contain the *neo$^r$* and *tk* genes and will be killed by gancyclovir. However, cells that underwent homologous recombination will contain the *neo$^r$* gene but lack the *tk* gene; these cells will therefore survive in the presence of both antibiotics (fig. 13.22). Geneticists can isolate embryonic stem cells from mice, cells that can produce any mouse tissue. The cells are transfected and then grown in tissue culture in the presence of neomycin and gancyclovir, and only cells that undergo homologous recombination will survive. These cells are then injected into early-stage mouse embryos to become part of the developing mouse. The mice that develop will be chimeric; some will have incorporated the transfected cells into the germ line and become heterozygotes for the disabled gene. Finally, when mice like this are mated, one fourth of their offspring will be homozygous for the disabled gene. Thus, knockout mice, have been created through this ingenious technique.

Knockout mice are especially useful for studying development and immunology. For example, if the gene for the Mullerian-inhibiting substance is knocked out, males are infertile because they develop female reproductive organs. This experiment led to insight into the genetic path for sex determination (see chapter 5). Hundreds of knockout experiments are published each year.

### Reporter Systems

We conclude this section by discussing two **reporter systems,** systems used to indicate that a transfection experiment was successful. Plants can be transfected with

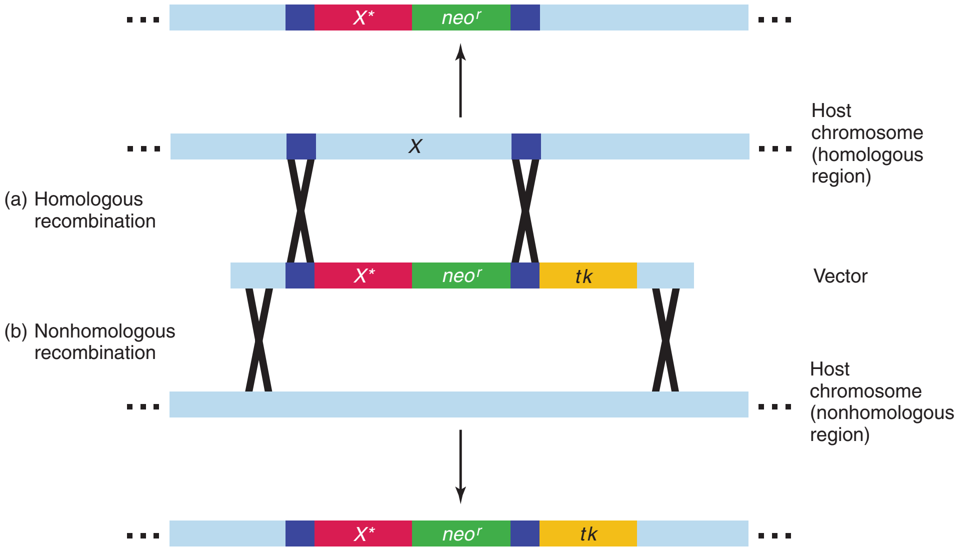

X*
neo$^r$
X
Host chromosome (homologous region)
(a) Homologous recombination
X*
neo$^r$
tk
Vector
(b) Nonhomologous recombination
Host chromosome (nonhomologous region)
X*
neo$^r$
tk

**Figure 13.22** Creating a knockout mouse. A vector is created that has a disabled (nonfunctional) form of the gene in question *($X^*$; red)*. Next to the gene in question is the neomycin resistance gene *(neo$^r$; green)*; both genes are surrounded by regions *(blue)* that flank the normal gene on its chromosome. Finally, outside the flanking regions in the vector is the gene for thymidine kinase *(tk; yellow)*. In homologous recombination (*a*), involving crossovers in the homologous flanking regions, the disabled gene and the neomycin resistance gene replace the normal gene on the cell's chromosome. In nonhomologous recombination (*b*), almost the entire vector is incorporated into the host chromosome, including the thymidine kinase gene. Techniques then allow for the selective growth of cells with the rare homologous recombination event. One fourth of the offspring of heterozygous chimeric mice will be knockouts for the gene in question.

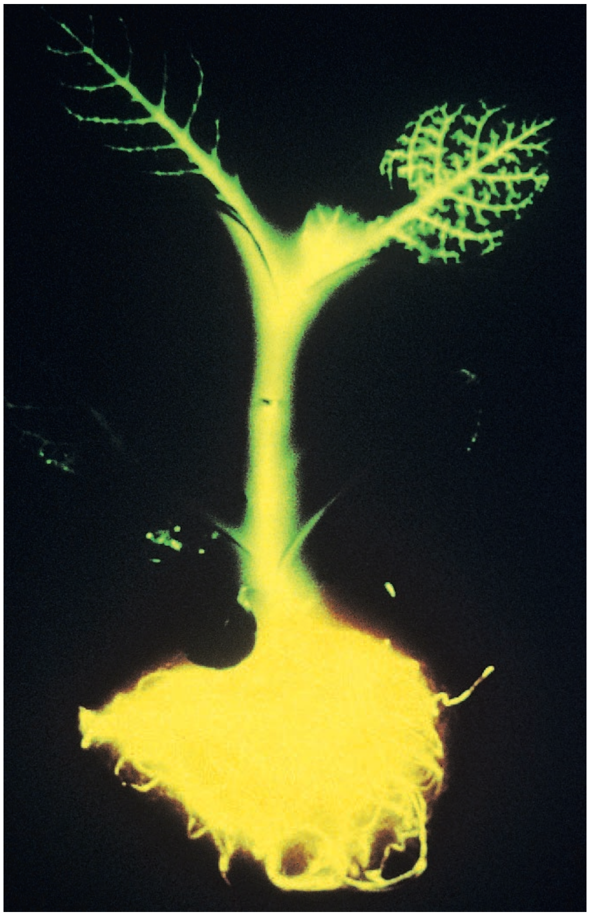


**Figure 13.23** Luminescent transgenic tobacco plant containing the firefly luciferase gene. The plant was watered with luciferin, resulting in a firefly glow. (© Science VU/Keith V. Wood/Visuals Unlimited.)

the Ti plasmids of *Agrobacterium tumefaciens,* as alluded to earlier. When a plant is infected with *A. tumefaciens* containing the Ti plasmid, a crown gall tumor is induced when the Ti plasmid transfects the host plant, transferring the T-DNA region. Those cells transfected with the T-DNA are induced to grow as well as to produce opines that the bacteria feed on. Much recent research has concentrated on engineering Ti plasmids to contain other genes that are also transferred to the host plants during infection, creating transgenic plants. One series of experiments has been especially fascinating.

Tobacco plants have been transfected by Ti plasmids containing the luciferase gene from fireflies. The product of this gene catalyzes the ATP-dependent oxidation of luciferin, which emits light. When a transfected plant is watered with luciferin, it glows like a firefly (fig. 13.23). The value of these experiments is not the production of glowing plants, but rather the use of the glow to "report" the action of specific genes. In further experiments, the promoters and enhancers of certain genes were attached to the luciferase gene. As a result, luciferase would only be produced when these promoters were activated; thus, the glowing areas of the plant show where the transfected gene is active.

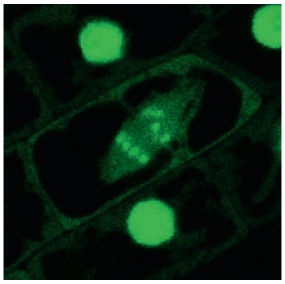


**Figure 13.24** Expression of green fluorescent protein in root cells of the plant *Arabidopsis thaliana* under fluorescent light. Chromosomes are visible in a cell undergoing mitosis and chromatin is visible as circles of green in interphase nuclei. The green fluorescent protein gene was fused to the carboxy terminus of the gene for the transcription factor Cry2, controlling genes for the phototrophic response (bending toward light). (Copyright Sean Cutler, Stanford University Plant Biology Department.)

One of the more recent reporter systems developed uses a gene from jellyfish that produces a **green fluorescent protein.** The value of this system is that it "reports" when ultraviolet light shines on it, rather than requiring an addition, as in the luciferase system. The gene for the green fluorescent protein is recombined with a gene in question, and then the transfection is performed. If the gene in question has transferred successfully, carrying the gene for the green fluorescent protein, the fluorescent protein will report it when activated by ultraviolet light (fig. 13.24).

## RESTRICTION MAPPING

The number of cuts that a restriction enzyme makes in a segment of double-stranded DNA depends on the size of that DNA, its sequence, and the number of base pairs in the recognition sequence of the particular enzyme. That is, a restriction enzyme with only three base pairs in its recognition sequence will cut more times than one with six base pairs in its sequence, since the probability of a sequence occurring by chance is a function of the length of that sequence. A sequence of three bases occurs more often by chance ($1/4^3 = 1/64$ base pairs) than a sequence of six bases ($1/4^6 = 1/4{,}096$ base pairs). *Hin*dII, for example, cuts the circular DNA of the tumor virus SV40 into eleven pieces; some restriction enzymes can cut *E. coli* DNA into hundreds of pieces. The product of the action of a restriction enzyme on a DNA sample is called a **restriction digest.**

Using electrophoresis, we can separate the fragments of a restriction digest by size. With techniques to be described later, we can locate the restriction sites on the original gene or piece of DNA. That is, we can construct a map of the restriction recognition sites that will give us the physical distance between sites, in base pairs (fig. 13.25). This **restriction map** is extremely valuable for several reasons. For example, when the radioactive nucleotide tritiated thymidine was added for a very short period of time during the beginning of DNA replication in SV40 viruses, the radioactivity always appeared in only one restriction fragment. This demonstrated that SV40 replication started from a single, unique point; that point was localized to a particular segment of the SV40 chromosome.

In addition, a restriction map often allows researchers to correlate the genetic map and the physical map of a chromosome. Certain physical changes in the DNA, such as deletions, insertions, or nucleotide changes at restriction sites, can be localized on the genetic map. These changes can be seen as changes in size, or in the total absence, of certain restriction fragments when compared with wild-type DNA. This information allows us to see changes in the DNA; it also gives us information about the evolution of species (see chapter 21). The differences in fragment sizes are called **restriction fragment length polymorphisms (RFLPs)** and have proven valuable in pinpointing the exact location of genes and determining the identity or relatedness of individuals. A restriction digest is also useful for isolating short segments of DNA that can be easily sequenced.

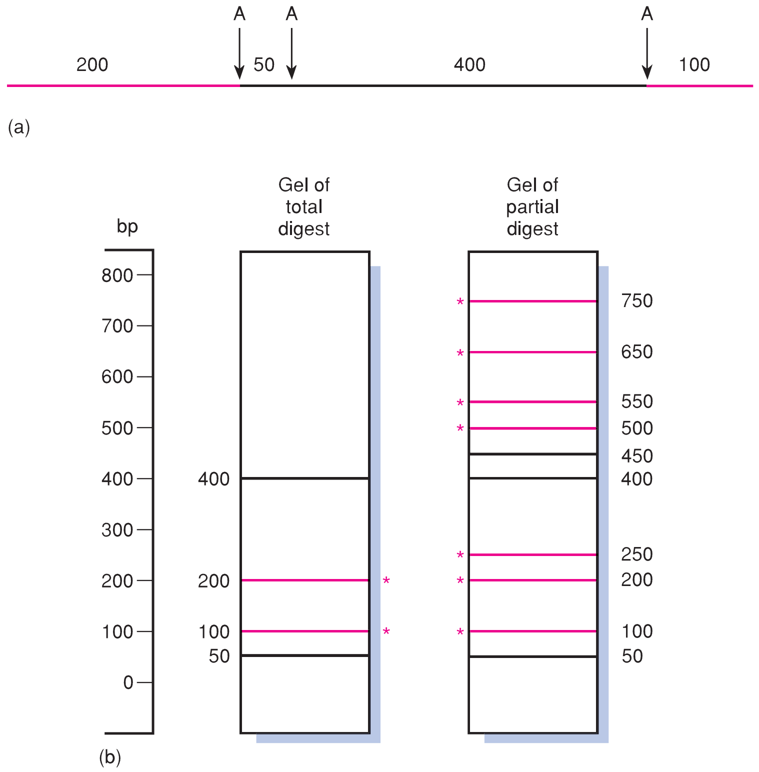


**Figure 13.25** Restriction map from electrophoresis of a restriction endonuclease digest. (*a*) Original piece of DNA, showing restriction sites marked by *A*. (*b*) Agarose gels showing bands of total and partial restriction digests. *Asterisks* mark radioactive bands produced by end-labeled segments. At the *left* is the scale of molecular weight markers in base pairs (e.g., 800 bp, 700 bp). The total digest produces fragments that are 400, 200, 100, and 50 bp—and the 200 and 100 bp fragments are end labeled. The partial digest yields six additional bands.

## Constructing a Restriction Map

How do we construct a restriction map? Figure 13.25 shows a hypothetical piece of DNA cut by restriction enzyme *A*. Below this map is a diagram of the electrophoresed digest on agarose gels, which are usually used because their porosity allows DNA fragments of relatively large size to move. The restriction enzyme makes three cuts in the DNA, generating four fragments that are 200, 50, 400, and 100 base-pairs long. The banding pattern on the gel at the *left* in figure 13.25*b* is the result of the electrophoresing of that digest. (Note that smaller segments move faster than larger segments.) The sizes of the segments are determined by comparison with standards of known size (not shown, although the scale is indicated on the left). The gel does not reveal the order of these segments on the chromosome. Several methods can be used to determine the exact order of the restriction segments on the original piece of DNA.

Before restriction enzyme digestion, the 5′ ends of the DNA can be labeled radioactively with $^{32}P$ using the enzyme polynucleotide kinase. Since the enzyme is acting on double-stranded DNA, both ends will be labeled. Upon electrophoresis after digestion of the DNA in figure 13.25, the 200-base-pair and 100-base-pair (bp) bands will be labeled radioactively, indicating that these segments are the termini of that piece of DNA. However, we still don't know the order of the middle pieces.

The order of the other segments can be determined by slowing down the digestion process to produce a **partial digest.** If the reaction is cooled or allowed to proceed for only a short time, not all restriction sites will be cut. Some pieces of DNA will not be cut at all, some will be cut once, some twice, and some cut at all three restriction sites. The result of electrophoresis of this partial digest is seen at the right in figure 13.25*b*. From this gel, we can reconstruct the segment order. This gel contains the four original segments plus six new segments, each containing at least one uncut restriction site.

From the total digest gel, we know that the 200 and 100 bp segments are on the outside because they were labeled radioactively. This means the 50 and 400 bp segments are on the inside. In the partial digest, we find a 250 bp segment but not a 150 bp segment, which tells us that the 50 bp segment lies just inside and next to the 200 bp terminus (fig. 13.26*b*). There is a 500 bp segment but not a 600 bp segment, which tells us that the 400 bp segment lies adjacent to the 100 bp terminus (fig. 13.26*c*). An unlabeled 450 bp segment confirms that the 400 and 50 bp segments are adjacent and internal in the DNA. We thus unequivocally reconstruct the original DNA (compare fig. 13.26*e* with fig. 13.25*a*), creating a map of sites of restriction enzyme recognition regions separated by known lengths of DNA.

## Double Digests

In practice, restriction mapping is usually done with several different restriction enzymes. Figure 13.27 is a map of the DNA of figure 13.25, with the recognition sites of a second endonuclease, *B,* included. Using the same methodology just outlined, we can show that the order of the *B* segments is 350, 250, and 150 base pairs arising from two cuts by endonuclease *B.* What we do not know is how to overlay the two maps. Do the *B* segments run left to right or right to left with respect to the *A* segments (fig. 13.27*a* and *b*)? We can determine the unequivocal order by digesting a sample of the original DNA with both enzymes simultaneously, thus producing a **double digest.**

The two orders shown in figure 13.27*a* and *b* are used to make different predictions about the double digest. From the first order (*a*), we predict a 200 bp end segment, radioactively labeled. From the second order (*b*), we predict that the labeled 200 bp segment will be cut back to 150 base pairs: there should not be a labeled 200 bp segment. The double digest shows a labeled 200 bp segment, indicating order (*a*). All other aspects of order (*a*) are consistent with the double digest.

Restriction mapping thus provides us with a physical map of a piece of DNA, showing restriction endonuclease sites separated by known lengths of DNA. This technique

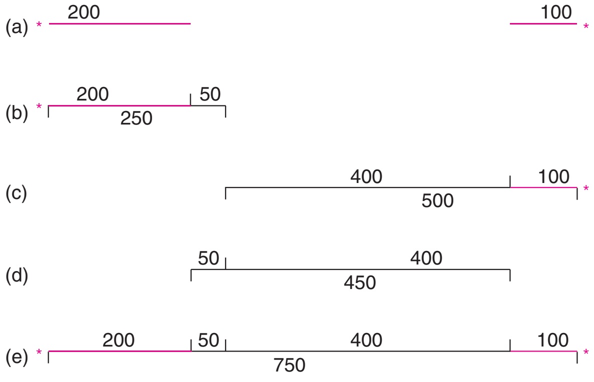


**Figure 13.26** Steps in the reconstruction of the DNA from figure 13.25. *Asterisks* show $^{32}P$ end labels. From the total digest, the 100 and 200 bp segments are established as the end segments (*a*). Since there are also 50 and 400 bp fragments within the DNA (established from the total digest), only certain bands (fragments) are possible from the partial digest, which establishes that the 50 bp fragment is adjacent to the 200 bp fragment and the 400 bp fragment is adjacent to the 100 bp end segment (steps *b* and *c*). The occurrence of an unlabeled 450 bp fragment in the partial digest verifies the existence of the 50 and 400 bp fragments (*d*), yielding the final structure (*e*). All the fragments in the partial digest are consistent with this arrangement.

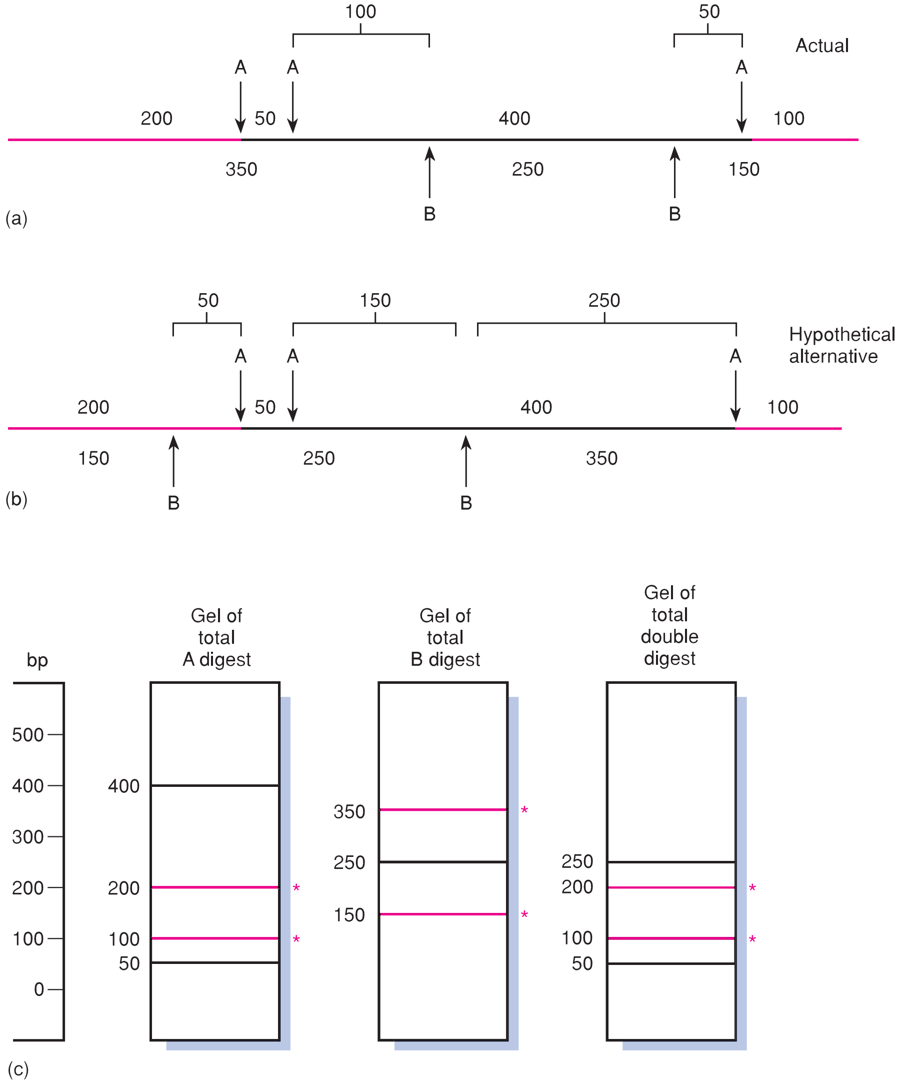

100
50
Actual
A
A
A
200
50
400
100
350
250
150
B
B
(a)
50
150
250
Hypothetical alternative
A
A
A
200
50
400
100
150
250
350
B
B
(b)
bp
500
400
300
200
100
0
Gel of total A digest
400
200
100
50
*
*
Gel of total B digest
350
250
150
*
*
Gel of total double digest
250
200
100
50
*
*
(c)

**Figure 13.27** Overlay of the recognition sites affected by two different restriction endonucleases (*A* and *B*) on the same piece of DNA. (*a*) Actual arrangement. (*b*) Hypothetical alternative arrangement. (*c*) Electrophoresis of the total restriction digests by A alone, B alone, and both. *Asterisks* indicate radioactive end-labeled bands. Order (*a*) is consistent with all the bands found in all the digests, whereas order (*b*) is not. For example, in order (*b*) an internal (unlabeled) 150 bp fragment is predicted, but this fragment is not found in the total digest.

gives us short DNA segments of known position that we can sequence, as well as a physical map of the DNA that can be compared with the genetic map and can locate mutations and other particular markers.

## *Restriction Fragment Length Polymorphisms*

Restriction fragment length polymorphisms (RFLPs), obtainable from restriction digests, are proving to be very valuable genetic markers in two areas of study: human gene mapping and forensics. In a restriction digest of the whole human genome, there might be thousands of fragments from a single restriction enzyme. Unique probes have been developed for Southern blotting these digests. Genetic variation usually comes in the form of a second allele that, due to a mutation, lacks a restriction site and is therefore part of a larger piece of DNA (fig. 13.28). Some probes have uncovered **hypervariable loci** with many alleles (any one

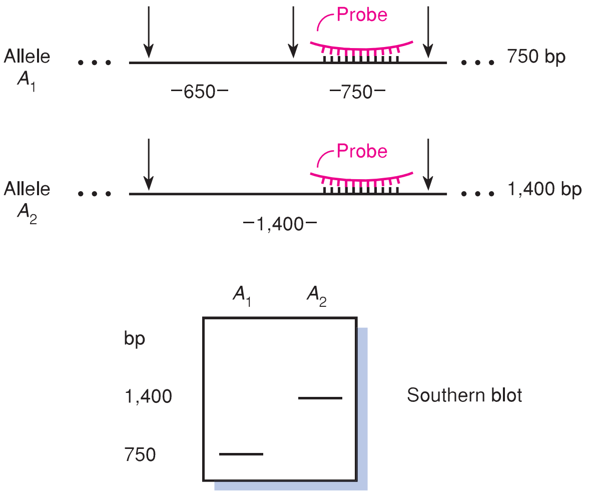


**Figure 13.28** Restriction fragment length polymorphism (RFLP) analysis. Allele $A_1$ is a gene segment, 750 bp long, identified by probe binding. In allele $A_2$, the restriction site to the *left* of allele $A_1$ has been changed. The probe thus recognizes a 1,400 bp fragment instead of the 750 bp fragment. During Southern blotting, two different bands show up from the two alleles. *Arrows* indicate restriction sites.

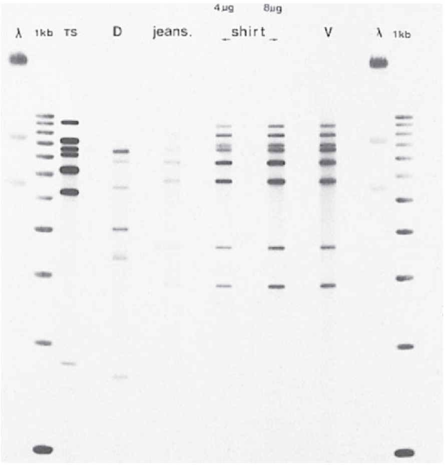


**Figure 13.29** Forensic use of DNA fingerprinting. Southern blot of DNA from victim (*V*) and defendant (*D*) in a crime. *Jeans* and *shirt* refer to blood samples taken from the clothing of the defendant. The pattern on the shirt clearly matches the victim's blood, not the defendant's own blood. All of the other lanes of the blot contain controls and size standards. The probability that the blood stains were not from the victim was estimated at one in thirty-three billion, more than the number of people on earth. However, these probabilities are controversial, depending on statistical assumptions about variability within racial and ethnic subpopulations. (Courtesy of Cellmark Diagnostics, Germantown, MD.)

person has, of course, only two of the many possible alleles). A population's genetic variation is generated because these hypervariable loci contain many tandem repeats of short (10 to 60 bp) segments. Due presumably to unequal crossing over (see chapter 8), just one of these loci, called **variable-number-of-tandem-repeats (VNTR) loci,** can generate much variation. As a result, probing for one of these VNTR loci in a population reveals many alleles.

The Southern blots of such digests create a **DNA fingerprint** of extreme value in forensics. DNA extracted from blood or semen samples left by a criminal can be compared with DNA patterns of suspects (fig. 13.29). When a single probe recognizes a number of different loci, each individual will have many bands on a Southern blot, with most people producing unique patterns. In one system, developed by A. Jeffreys, a single probe locates fifty or more variable bands per person. If Jeffreys's probes are used to compare the patterns, the likelihood that the two patterns would match randomly is infinitesimally small. This technique thus has greater power to identify individuals than using the prints from their fingertips.

Alec Jeffreys (1950– ). (Courtesy of Dr. Alec Jeffreys.)

## POLYMERASE CHAIN REACTION

In the past, in many instances (in museum specimens, dried specimens, crime scene evidence, and fossils), a DNA sample was available, but in such small quantity or so old as to be considered useless for study. That situation changed in 1983 when Kary Mullis, a biochemist working for the Cetus Corporation, devised the technique we now refer to as the **polymerase chain reaction (PCR).** PCR can be used to amplify whatever DNA is present, however small in quantity or poor in quality. The only requirement is that the sequence of nucleotides on either side of the sequence of interest be known. That information is needed to construct primers on either side of the sequence of interest. Once that is done, the sequence between the primers can be amplified.

In the PCR technique, the primers and the ingredients for DNA replication are added to the sample. Then, the mixture is heated (e.g., 95° C for twenty seconds) to denature the DNA. The temperature is then lowered (e.g., 55° C for twenty seconds) so that primers can anneal to their complementary sequences. The temperature is then raised again (e.g., 72° C for twenty seconds) for DNA replication. Then, a new cycle of replication is initiated (fig. 13.30). The various stages in the cycle are controlled by changes in temperature since the temperatures for denaturation, primer annealing, and DNA replication are different. About twenty cycles of PCR produces a million copies of DNA; thirty cycles make a billion copies. The technique is aided by using DNA polymerase from a hot-springs bacterium, *Thermus aquaticus,* that can withstand the denaturing temperatures. Thus, after each cycle of replication, no new components have to be added to the reaction mixture. Rather, the cycling can be continued without interruption in PCR machines (simply programmable water baths that accurately and rapidly change the water temperature that surrounds the reaction mixture). Some machines can process ninety-six samples at a time.

PCR has been used to create DNA fingerprints by amplifying **microsatellite DNA.** These are repeats of very short sequences of DNA dispersed throughout the genome. For example, cytosine-adenine (CA) repeats occur tens of thousands of times in eukaryotes, in repeats of from twenty to sixty base pairs. As in the case of VNTR loci, there is tremendous variability among people in the number of these repeats at a locus, due presumably to crossover errors. Unlike the situation with VNTR loci, however, PCR amplification of one of these loci can be done without restriction cutting, Southern blotting, and probing—PCR gives the results directly upon electrophoresis. All we need are the surrounding primer sequences to any microsatellite locus. PCR is now a routinely used tool in the laboratories of molecular geneticists. They use it to rapidly amplify the DNA regions of interest for research or forensic uses.

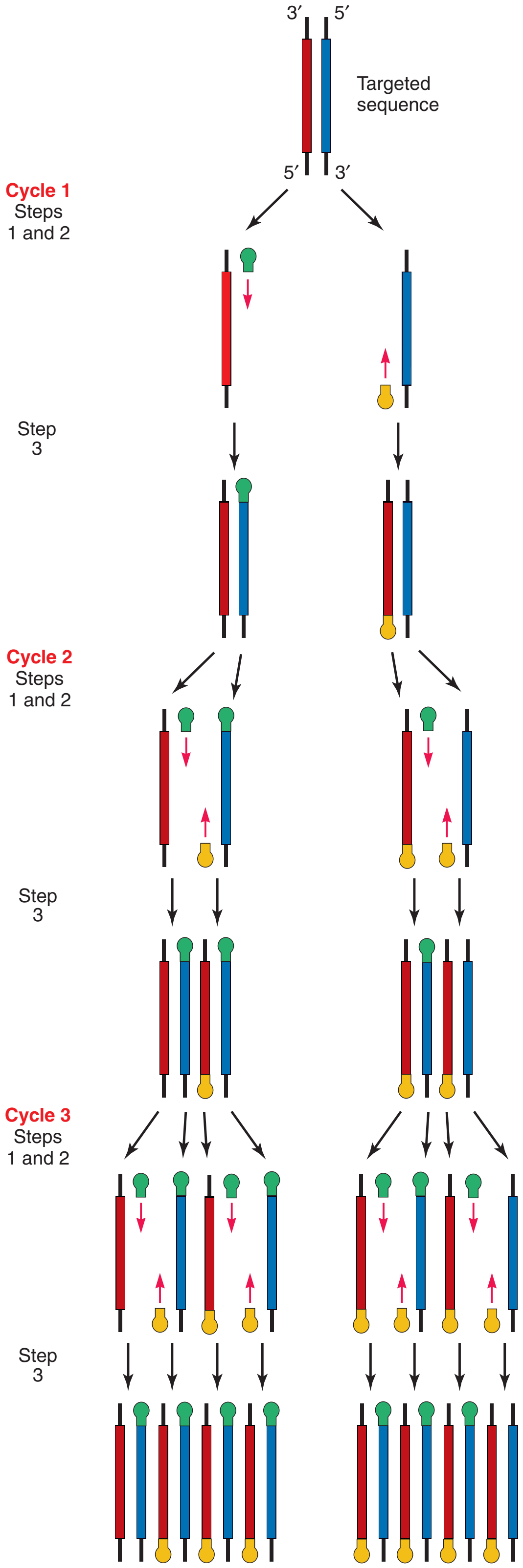


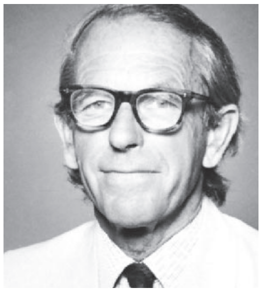
**Figure 13.30** Polymerase chain reaction. DNA is denatured, (step 1), primer oligonucleotides that are complementary to end sequences on the two strands anneal (step 2), and DNA replication takes place (step 3). Each step in the cycle is controlled by temperature changes. The targeted sequence is shown as *red* on one stand and *blue* on the other. Primers are shown as either *green* or *yellow* lollipops. A green primer begins the copying of the red strand into a complementary blue strand; a yellow primer begins the copying of a blue strand into a complementary red stand. In three cycles, one double-stranded region of DNA becomes eight. The process requires the addition of primers, deoxynucleotide triphosphates, and DNA polymerase, as well as changing temperature cycles.

We now turn our attention to a major result of recombinant DNA technology, DNA sequencing. Recombinant DNA technology, with its ability to isolate and amplify small, well-defined regions of chromosomes, has allowed the development of DNA sequencing techniques.

## DNA SEQUENCING

Paul Berg of Stanford University, Walter Gilbert of Harvard University, and Frederick Sanger of the Medical Research Council in Cambridge, England, shared the 1980 Nobel Prize in chemistry. Berg won for creating the first cloned DNA molecules when he spliced the SV40 genome into phage λ. Gilbert and Sanger were awarded the prize for independently developing methods of sequencing DNA. Gilbert, along with Allan Maxam, developed a method of DNA sequencing called the *chemical method.* It involves chemically breaking down the DNA at specific bases. Sanger, who won a Nobel Prize in 1959 for sequencing the insulin protein, later took part in developing methods for sequencing RNA. His sequencing method, developed with Alan Coulson, involved DNA synthesis and was called the *plus-and-minus method.* The further development of the method by Sanger, Coulson, and S. Nicklen, using specific chain-terminating nucleotides, led to a modification of the plus-and-minus method known as the *dideoxy method.*

Walter Gilbert (1932– ).
(Photo: Rick Stafford.)

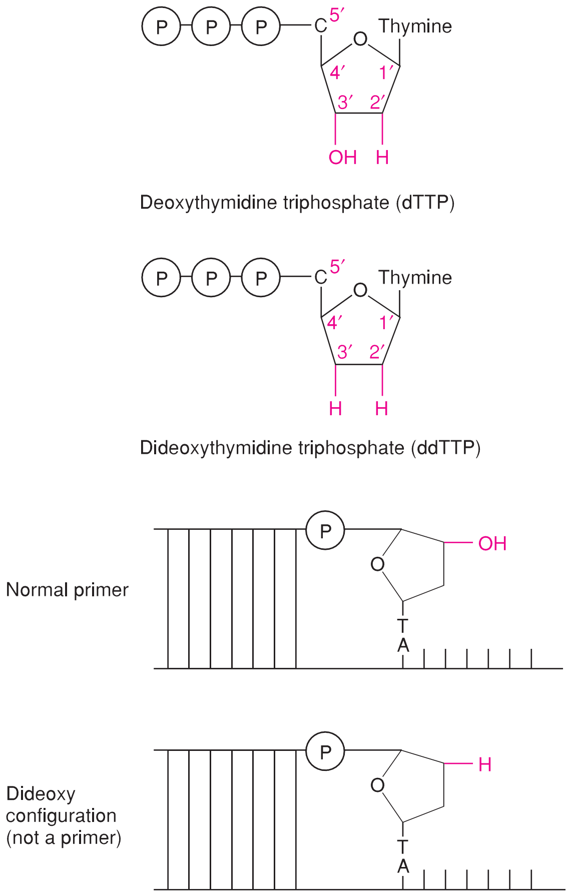


Frederick Sanger (1918– ).
(Courtesy of Dr. Frederick Sanger.)

### *The Dideoxy Method*

In the **dideoxy method,** manipulation of DNA synthesis enables DNA sequencing. Remember from chapter 9 that DNA synthesis occurs at a primer configuration, one in which double-stranded DNA ends with a 3′-OH group on one strand. The other strand continues as single-stranded DNA (fig. 13.31, *middle*). The dideoxy method creates a primer configuration of the DNA to be sequenced and enables replication to proceed. A trick, using chain-terminating nucleotides, stops DNA synthesis at known positions. These chain-terminating nucleotides are formed of sugars lacking OH groups at both the 2′ and 3′ carbons (hence the term *dideoxy*). Without a 3′-OH group, a dideoxynucleotide cannot be used for further DNA polymerization (fig. 13.31).

Chain-terminating nucleotides permit synthesis to be stopped at a known base. The sample to be sequenced is elongated separately in four different reaction mixtures, each having all four normal nucleotides but also having a proportion of one of the chain-terminating dideoxy nucleotides. For example, if the pool of thymine-containing triphosphate nucleotides contains a portion of the dideoxythymidine triphosphate molecules, then synthesis of the growing strand is sometimes terminated when adenine (the complement of thymine) appears on the template, creating fragments that end in thymine. Similar



Deoxythymidine triphosphate (dTTP)

Dideoxythymidine triphosphate (ddTTP)

**Figure 13.31** Dideoxy nucleotides cause chain termination during DNA replication. The dideoxy primer configuration lacks the 3′-OH group needed for chain lengthening in a normal primer configuration.

reactions are carried out in separate test tubes for each of the other nucleotides, producing fragments that terminate when the respective complementary nucleotide is present. The resulting fragments from each reaction are electrophoresed, generating a pattern on the gel that reveals the sequence of the newly synthesized DNA. Let us go through an example.

In figure 13.32*a*, we show the DNA to be sequenced, a small segment of nine base pairs. To sequence this segment, one must get one strand of this double-stranded segment into the configuration shown in figure 13.32*b*. The DNA to be sequenced must be the template for new DNA synthesis. (We will soon discuss how we obtain the required configuration.) Having created the necessary primer configuration, we take four subsamples of it, each including all four nucleoside triphosphates plus DNA polymerase I. At least one of the nucleoside triphosphates is radioactively labeled, usually with $^{32}$P. This label allows us to identify newly synthesized DNA by autoradiography.

To each of the four subsamples, one of the dideoxynucleotides (dd) is added—one subsample gets ddTTP, one gets ddATP, one gets ddCTP, and one gets ddGTP. These dideoxynucleotides are added in addition to the regular deoxynucleotides to increase the probabil-

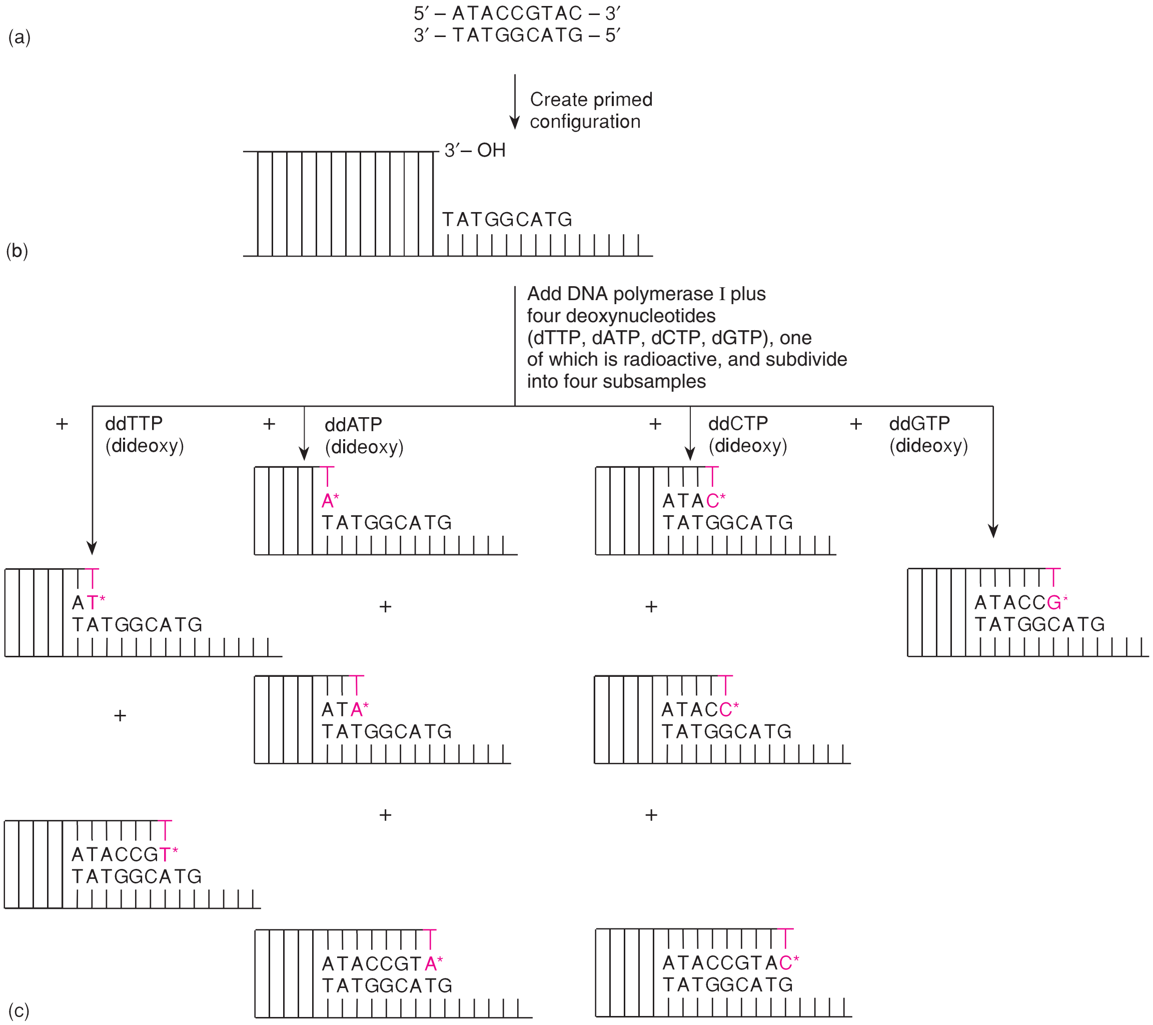

(a)
5′ – ATACCGTAC – 3′
3′ – TATGGCATG – 5′
Create primed configuration
3′– OH
TATGGCATG
(b)
Add DNA polymerase I plus four deoxynucleotides (dTTP, dATP, dCTP, dGTP), one of which is radioactive, and subdivide into four subsamples
+ ddTTP (dideoxy)
+ ddATP (dideoxy)
+ ddCTP (dideoxy)
+ ddGTP (dideoxy)
AT*
TATGGCATG
+
ATACCGT*
TATGGCATG
A*
TATGGCATG
+
ATA*
TATGGCATG
+
ATACCGTA*
TATGGCATG
ATAC*
TATGGCATG
+
ATACC*
TATGGCATG
+
ATACCGTAC*
TATGGCATG
ATACCG*
TATGGCATG
(c)

**Figure 13.32** Initial steps in the dideoxy method of DNA sequencing. The *asterisks* indicate the dideoxynucleotides. The DNA to be sequenced is placed into a primer configuration (*a*, *b*). Four reaction mixtures are created, each with all four normal nucleotides plus one of the dideoxynucleotides. Thus DNA synthesis in each reaction mixture is stopped a percentage of the time when the complement to the dideoxynucleotide appears in the template (*c*).

ity that chain termination will occur at every appropriate position. If the dideoxynucleotide were added in place of the deoxynucleotide, then the chain would be terminated the first time the complement of that base appeared in the template strand. By mixing the dideoxynucleotides and the deoxynucleotides, we are assured that termination will occur in every appropriate position.

In figure 13.32*c,* we see that the template has two adenines. Therefore, in the ddTTP reaction mixture, adenine's complement (thymine) is needed twice. There are thus two possible points for ddTTP to incorporate, two possible chain terminations, and therefore two fragments that could end in dideoxythymidine, of two and seven bases, respectively. Similarly, there are three possible fragments ending in adenine, of one, three, and eight bases; three ending in cytosine, of four, five, and nine bases; and one ending in guanine, of six bases (fig. 13.32*c* and fig. 13.33, *top*).

After DNA synthesis is completed, the old primer is removed, leaving only newly synthesized DNA fragments (fig. 13.33). Newly replicated segments of various lengths

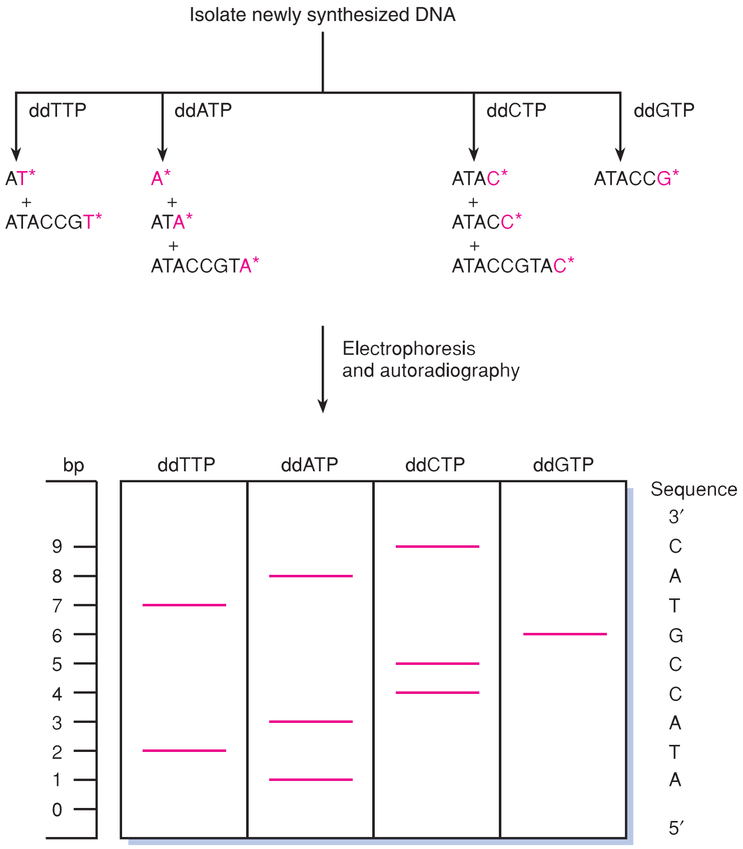


**Figure 13.33** Electrophoresis of segments produced by the dideoxy method of DNA sequencing. This method allows direct reading of the sequence. The *asterisks* indicate the dideoxynucleotides. The newly synthesized reaction products seen in figure 13.32 are isolated by removal of the primer and template. Each reaction mixture (e.g., ddTTP is the mixture containing dideoxythymidine triphosphates) produces specific products of specific lengths that can be determined by electrophoresis. In the case of the ddTTP mixture, two fragments ending in thymine are possible; one is two bases long, the other seven bases long. Thus, the complement of thymine, adenine, appears in positions 2 and 7 of the original piece of DNA. However, either the original strand or its complement (the new synthesis) yields the original sequence since DNA is a double helix; the sequence in one strand is always defined by the complementary sequence in the other strand.

from each reaction mixture are placed in separate slots and then electrophoresed on polyacrylamide gels to determine the lengths of the segments. Since only newly synthesized DNA segments are radioactive, autoradiography lets us keep track of newly synthesized DNA. As you can see from the autoradiograph of the gel in figure 13.33, each subsample produces segments that begin at the primer configuration (beginning of synthesis) and end with the chain-terminating dideoxy base. By starting at the bottom and reading up, back and forth across the gel, we can directly determine the exact sequence of the DNA segment. Because they have the appearance of stepladders in each lane (fig. 13.34), the gels are usually referred to as **stepladder gels** or **ladder gels.**

This technique (in the form of the original plus-and-minus method) was first used to sequence the genome of the DNA phage ϕX174 (box 13.3). That phage was used because it lent itself to the sequencing method. It has single-stranded DNA within the phage coat, yet its DNA becomes double-stranded once it enters the bacterium. Creating a primer configuration was thus relatively easy. The double-stranded circle from within the host could be treated with a restriction endonuclease to produce double-stranded fragments (fig. 13.35). These fragments could then be denatured. From this mixture, a particular fragment could be isolated by electrophoresis. The isolated strand would reanneal to the single-stranded DNA taken from the phage heads, forming a primer for new growth. The same restriction endonuclease would free the new growth after it had taken place. Thus, the dideoxy method was relatively easy to apply to the 5,387-base chromosome of ϕX174.

## Creating a General-Purpose Primer

To make the dideoxy method efficient, researchers created a general primer for routine sequencing work by recombinant DNA engineering of an *E. coli* vector, the single-stranded DNA phage M13. This phage is similar to ϕX174 in that both are packaged as single-stranded DNA, and both are replicated to double helices within the host. Therefore, the double-stranded form within the host, called the replicating form, can be engineered by standard methods, and the single-stranded form can be used for sequencing. The system works as follows.

By very clever engineering, J. Messing and his colleagues created cloning sites for a variety of restriction enzymes in a bacterial gene (*lacZ*) that had been inserted into M13 (fig. 13.36). The gene is for the β-galactosidase enzyme that normally breaks down lactose. It also breaks down an artificial substrate of the enzyme, X-gal, which is normally colorless. When cleaved by β-galactosidase, X-gal becomes blue. Thus, in the presence of the functional *lacZ* gene, M13 plaques are blue. If the gene is disrupted by a cloned insert, X-gal does not break down,

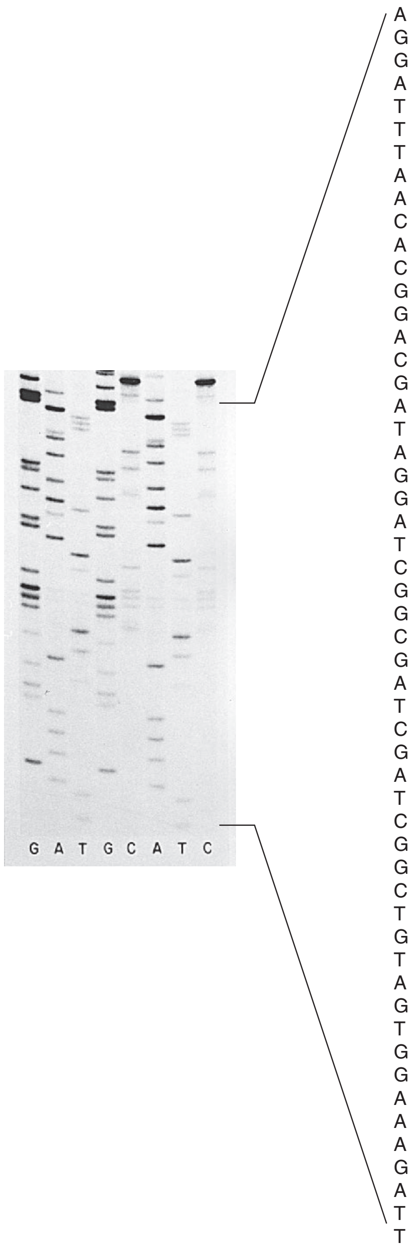


**Figure 13.34** Autoradiograph of a dideoxy sequencing gel. The letters *G, A, T,* and *C* along the *bottom* refer to the ddGTP, ddATP, ddTTP, and ddCTP reaction mixtures, respectively. Lanes are repeated for easier identification of the bands. The sequencing is also verified by sequencing the complementary strand and checking for agreement. (Courtesy of Richard J. Roberts.)

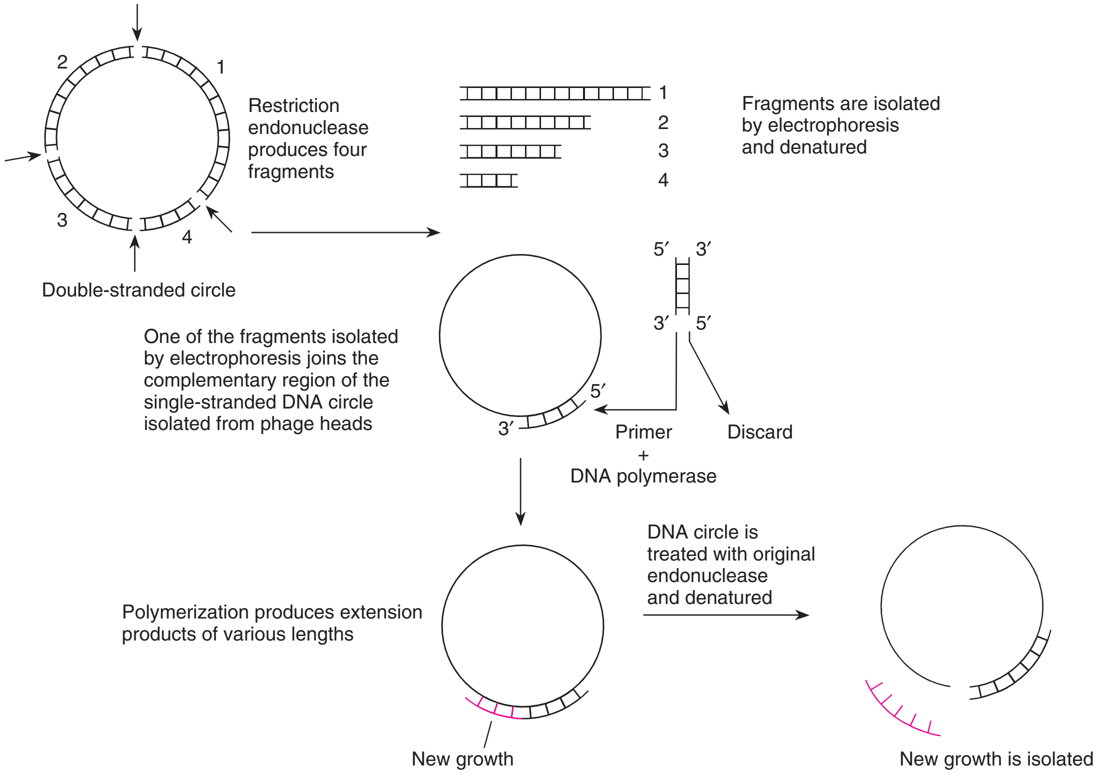

2
1
Restriction endonuclease produces four fragments
3
4
Double-stranded circle
1
2
3
4
Fragments are isolated by electrophoresis and denatured
5′
3′
3′
5′
One of the fragments isolated by electrophoresis joins the complementary region of the single-stranded DNA circle isolated from phage heads
5′
3′
Primer + DNA polymerase
Discard
DNA circle is treated with original endonuclease and denatured
Polymerization produces extension products of various lengths
New growth
New growth is isolated

**Figure 13.35** The genome of phage ϕX174 lent itself to the dideoxy method (originally, the plus-and-minus method) of DNA sequencing. Because the phage occurs in both the single- and double-stranded forms, it can be manipulated for sequencing. The double-stranded form is fragmented with an endonuclease. One fragment is isolated by electrophoresis and hybridized to the single-stranded form, creating a primer for new DNA synthesis and thus for dideoxy sequencing. Newly synthesized DNA can be isolated by treating it with the same restriction enzyme, which will create the same cut made originally. The newly isolated pieces can then be electrophoresed as in figure 13.34.

and hence the plaques are colorless. (M13 doesn't form true plaques because it doesn't lyse the *E. coli* cells. It does form turbid sites due to reduced bacterial growth.)

An oligonucleotide primer can be synthesized that is complementary to a region of the phage DNA upstream from the cloning sites. Single-stranded phage DNA containing a cloned insert is isolated and hybridized with the synthetic oligonucleotide. This operation creates the primer configuration for dideoxy sequencing of the cloned DNA. Virtually any clonable segment of DNA can be sequenced using this very general method. Theoretically, that segment could be any size.

Stepladder gels, however, are effective only up to about four hundred base pairs. To sequence larger regions requires sequencing overlapping segments and reconstituting the sequence by the overlap pattern, similar to the methods we described for amino acid sequencing (chapter 11, box 11.1). Overlapping segments of DNA are usually obtained by using two or more restriction enzymes.

The most recent innovation in DNA sequencing involves using four fluorescent dyes, each fluorescing at a different wavelength (505, 512, 519, and 526 nm); each of the four dideoxy nucleotides has a different dye attached. After the newly synthesized fragments are isolated, the products from all four reactions are run together in the same lane of a polyacrylamide gel. The gel is then scanned with an argon laser that excites the dye molecules. An instrument records the color of the peaks, reading the sequence directly and automatically (fig. 13.37). This method greatly simplifies sequencing since it is automated. It also alleviates the necessity for radioactive tags.

## BOX 13.3

## Experimental Methods

### *Genes Within Genes*

Complete sequencing of a DNA genome using Sanger and Coulson's plus-and-minus method (the forerunner to the dideoxy method) was first accomplished with ϕX174, a virus that contains a single-stranded DNA circle of 5,387 bases within its protein capsule. Once injected into the host, the DNA is replicated to form a double helix that then proceeds in normal viral fashion to replicate itself, manufacture its own coat proteins, lyse the cell, and escape. This virus has nine genes. The virion is a small, twenty-faced polyhedron with a small spike at each of its twelve vertices. This spike attaches ϕX174 to *E. coli.* The coat accounts for one protein and the spike accounts for two. Thus, three of the virus's nine genes manufacture coat proteins. Figure 1 illustrates the location of the genes in ϕX174, obtained through standard mapping methods.

From the information obtained from the sequencing of MS2, an RNA virus, geneticists believed that there should always be a nontranslated sequence between genes, presumably for the purpose of controlling expression of each gene. However, careful perusal of the nucleotide sequence of ϕX174 provided several surprises. First, the ends of three genes overlapped the beginnings of the next genes (*A-C, C-D,* and *D-J*); in the first two cases, the initiation codon is entirely within the end of the previous gene, but read in a different frame of reference. In the sequence ATGA, the ATG is the initiation of the next gene, whereas the TGA is the termination of the previous gene. In the *D-J* interface, one A is shared: TAATG (UAAUG in ribose nucleotides; fig. 2). It is the number 3 base of the termination codon and the number 1 base of the initiation codon. The surprises did not end there.

At first, with the sequence of nucleotides spread out in front of them, the researchers could not find the *B* and the *E* genes; they appeared to be missing. Upon careful analysis, however, the scientists found that the *B* gene was entirely within the *A* gene and the *E* gene was entirely within the *D* gene (fig. 3). Their finding went against theory. We were led to believe, from logical arguments, that genes cannot substantially overlap. There would be too much of a constraint on function: The functional sequence of one gene would also have to be a functional sequence in the other. Similarly, there would be an evolutionary constraint involved. The genes would have to evolve together. But here we have two cases in which genes do overlap. How could overlapping genes come about?

There are a large number of thymine bases in the ϕX174 genome. In the *D* gene particularly, many of the codons end with thymine. The imbedded *E* gene is read on a shifted frame with *D* so that the terminal bases of *D*'s codons are the middle bases of *E*'s. A look at the genetic code (see table 11.4) shows that the codons with U in the middle (*E*'s codons) are mainly for hydrophobic amino acids. Thus, *E* is a protein with detergent properties. In fact, it is the protein responsible for the dissolution of the outer cell wall of the host bacterium, a process that a detergent can accomplish in vitro. The properties of the *E* gene, then, are more the properties of its individual amino acids rather than their exact sequence.

In the *A-B* case, there is an indication that the two genes were once autonomous. This indication is based on the patterns of the codons; *A*'s codons tend to end in thymine before the overlap, but thereafter, in the region of overlap, *B*'s codons end in thymine, whereas *A*'s codons do not. Presumably, a mutational event tagged the *B* material onto the end of the earlier, shorter *A* gene and improved its enzymatic ability. We can only speculate, however.

The amazing arrangement of this viral DNA is one of extreme economy. The protein package is small, yet a minimum of nine genes had to be packed into it. We have seen this kind of economy before in the codon usage of mitochondrial DNA (see chapter 11).

As more sequencing has taken place, geneticists have discovered other novel overlap situations. For example, in one case, two genes were transcribed from opposite strands of the same region of DNA from a rat. On one strand, the gonadotropin-releasing hormone gene (*GnRH*) is located. On the other is a gene (*RH*) that produces a protein expressed in the heart.

Gene overlap is known to occur in bacteria as well. In *E. coli,* the promoter for the *ampC* gene (coding for the enzyme β-lactamase) begins within the last ten codons for the *frdC* gene, which codes for a subunit of the enzyme fumarate reductase. There is evidence that in this arrangement the *frdC* terminator can have some regulatory control of *ampC* transcription. (See chapter 14 for a discussion of regulatory processes in prokaryotes.)

With DNA sequence data, including the complete sequences of other chromosomes such as those of SV40 and mitochondria, we have accumulated much information about gene arrangements. Overlap to one degree or another has been found in small viruses (ϕX174, SV40), large viruses (λ), mitochondrial chromosomes, bacterial DNA, and even eukaryotes, in which several cases are now known in which genes are located within introns of other genes. In one of the few examples known, three genes are located in an intron of the neurofibromatosis gene, a gene that causes a disfiguring neurological disease. Although relatively uncommon, overlap and embedding of genes may have some regulatory role in transcription in addition to minimizing the length of the chromosome.

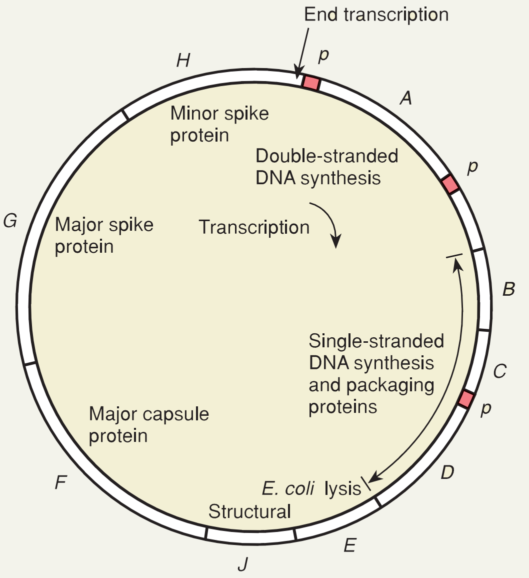

End transcription
p
H
A
Minor spike protein
Double-stranded DNA synthesis
p
G
Major spike protein
Transcription
B
Single-stranded DNA synthesis and packaging proteins
C
p
Major capsule protein
*E. coli* lysis
D
F
Structural
E
J

**Figure 1** Presumed location of the nine genes of phage ϕX174 on its circular chromosome. Transcription begins at three different places, each marked *p*, for promoter. The function of each gene appears within the circle.

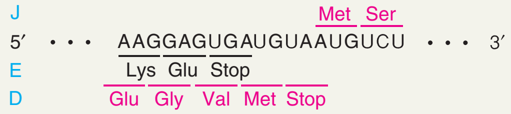

J
Met Ser
5′ • • • AAGGAGUGAUGUAAUGUCU • • • 3′
E
Lys Glu Stop
D
Glu Gly Val Met Stop

**Figure 2** Sequence, shown as ribose nucleotides, where genes *E* and *D* end and gene *J* begins. Each is out of register with the other two. The A of AUG for gene *J,* for example, is the second A of the UAA terminator of gene *D*.

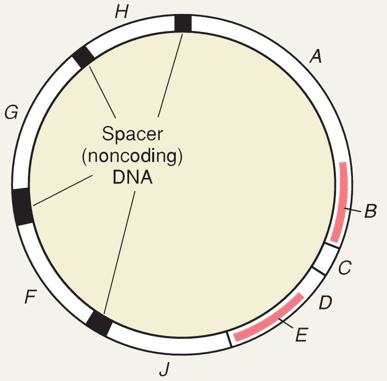

H
A
G
Spacer (noncoding) DNA
B
C
F
D
E
J

**Figure 3** The actual map of the nine genes of phage ϕX174. Note that *B* is entirely within *A* and *E* is entirely within *D.*

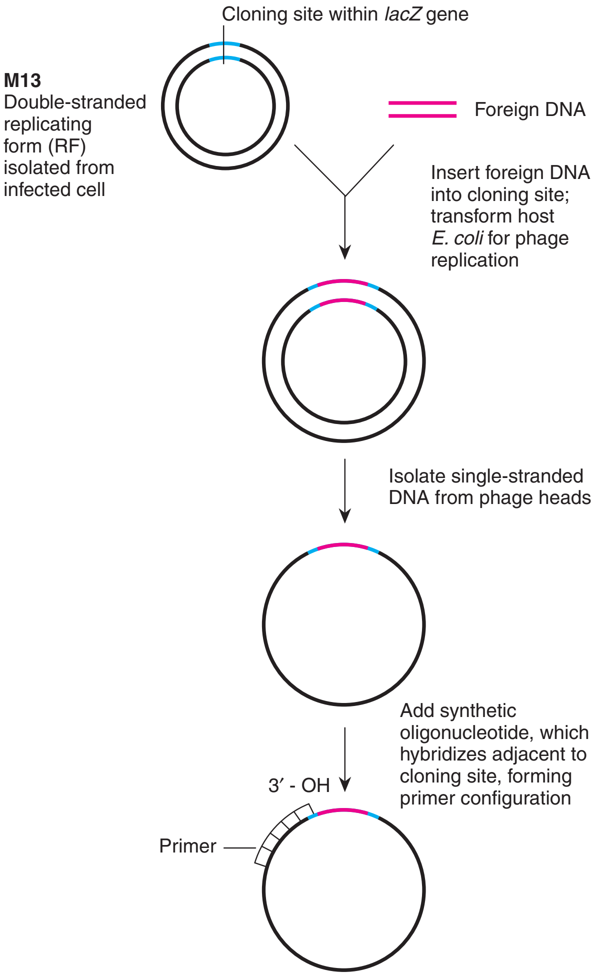

Cloning site within *lacZ* gene
**M13**
Double-stranded replicating form (RF) isolated from infected cell
Foreign DNA
Insert foreign DNA into cloning site; transform host *E. coli* for phage replication
Isolate single-stranded DNA from phage heads
Add synthetic oligonucleotide, which hybridizes adjacent to cloning site, forming primer configuration
3′ - OH
Primer

**Figure 13.36** Phage M13, a useful vector for sequencing a piece of cloned DNA by the dideoxy method since it exists in both single- and double-stranded forms. In addition, it contains restriction sites within a copy of the *lacZ* gene (*blue*). This allows for the selection of clones with inserted pieces of foreign DNA. An artificial oligonucleotide, which hybridizes adjacent to the cloning site, provides the primer configuration needed for new synthesis.

Either the dideoxy or the chemical method of sequencing (not discussed) allows us to read the sequence of hundreds of nucleotides on a single gel. Whole viral, prokaryotic, and eukaryotic genomes, and numerous regions of interest in prokaryotes, eukaryotes, and viruses have been sequenced. As W. Gilbert said in his Nobel Prize acceptance speech in 1981, "When we work out the structure of DNA molecules, we examine the fundamental level that underlies all processes in living cells."

## MAPPING AND SEQUENCING THE HUMAN GENOME

### *Locating a Gene of Interest*

Genes of importance can be searched for directly. A breast cancer gene provides a good example. Other genes that have been found this way include the genes for cystic fibrosis and Huntington disease. The concept of finding a gene is relatively simple; the methodology is tedious. Searching for many genes, including medically important genes such as one for breast cancer, means looking for a gene only by its symptoms; that is, we don't know the protein product of the gene or its location. Searching begins by looking at pedigrees of families segregating the disease and then trying to correlate the occurrence of the disease with a particular RFLP or microsatellite marker. When this is done, the gene has been localized to a particular region of a particular chromosome. Then, with a genomic library, chromosome walking (see the next section) is done until a gene in the neighborhood of the marker is found that could be the target gene. With the gene in hand, its sequence and protein product can be determined, a first step in medical treatment.

### *Chromosome Walking*

Despite the limited size of any one inserted piece of foreign DNA, it is possible to learn about longer stretches of DNA by using a technique of overlapping clones called **chromosome walking.** Let us say that a particular gene (in region A) is located in clone 1, as discovered through probing. The cloned insert can be removed, using the same restriction enzyme initially used to insert it in the vector, and broken into small pieces that are used as probes themselves. The idea is to locate another clone with an inserted region that overlaps the first one (fig. 13.38). The second clone is now treated the same way—with segments used to probe for yet another overlap farther down the chromosome. In this way, relatively long segments of a chromosome can be available for study in overlapping clones.

One obvious use of chromosome walking is to discover what genes lie next to each other on eukaryotic chromosomes. The technique is very tedious and is halted at certain areas not amenable to walking, such as repeated sequences found in the DNA of eukaryotes (see chapter 15). Once an overlapping probe contains a commonly repeated sequence, it hybridizes to many clones that do not contain adjacent segments. This "cross-referencing" lessens the value of the technique. Currently, newer techniques (termed **chromosome jumping**), designed to bypass regions not amenable to walking, are being developed. These techniques depend on the ability to locate

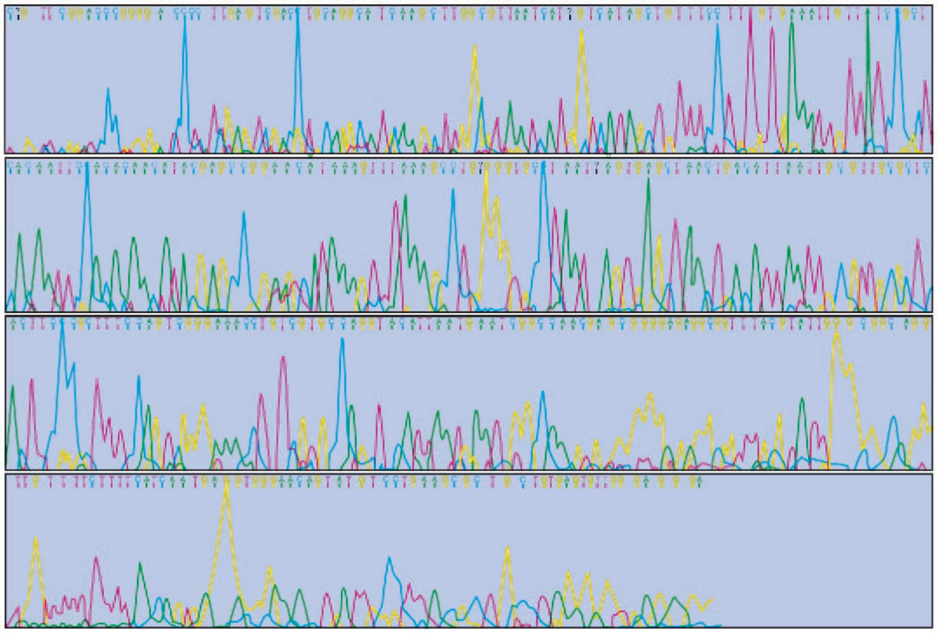
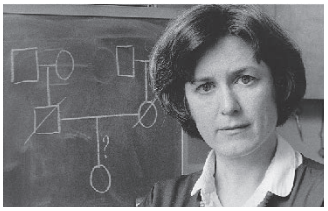


**Figure 13.37** Processed data from automated DNA analysis using fluorescent dyes. The DNA is sequenced by attaching a different fluorescent dye to each dideoxy base. Thus, the dideoxy bases can be identified by their fluorescent color in a laser light rather than by which lane they occupy in a gel. Only one lane need therefore be run. In this diagram, guanine is *yellow,* cytosine is *blue,* adenine is *green,* and thymine is *red.* The sequence is read *left to right, top to bottom.* (From L. Johnston-Dow, et al., *BioTechniques,* 5:754–65, 1987, copyright © 1987. Eaton Publishing, Natick, MA. Reprinted with permission.)

the two ends of a segment without having to walk through the middle. Ends of a segment can be located if the region has been inverted or if a large region is cloned and the middle part later removed, leaving just the ends. A probe of the ends allows the investigator to locate clones with first one end and then the other, effectively jumping over the intervening region.

### *The Breast Cancer Gene*

The initial location of the breast cancer gene *BRCA1* was determined by M. King in 1990 using a marker (*D17S74*) on the long arm of chromosome 17 (fig. 13.39); it was the 183rd marker that King had tried (fig. 13.40). The breast cancer gene *BRCA1* was particularly difficult to locate because it accounts for only about 5% of all breast cancers. However, it accounts for a much higher percentage of inherited, early onset breast cancers, those in women under fifty years of age. One woman in two hundred inherits this gene, and among those women, 80 to 90% risk developing the disease. The actual locating and cloning of this gene was done in 1994 by a team led by M. Skolnick. The gene codes for a protein of 1,863 amino acids; it seems to act as a tumor suppressor protein (see chapter 16). Its mechanism of action is as a transcription factor associated with RNA polymerase II (see chapter 10).



Mary-Claire King (1946– ). (Courtesy of Office of Public Information, Berkeley Campus, University of California. Photograph © Jane Scherr.)

## The Human Genome Project

### *The Standard Method*

In chapter 6, we developed a human chromosome map. Generally, a locus was located on a particular chromosome by tissue culture techniques (somatic-cell hybridization). Loci could be pinpointed further using aberrant chromosomes, such as those with deletions. If a locus was present when the intact human chromosome was present but absent if the deletion chromosome was present, the gene could be localized to the deleted region. In addition, probes for specific genes can show us roughly where that gene is located (fig. 13.41).

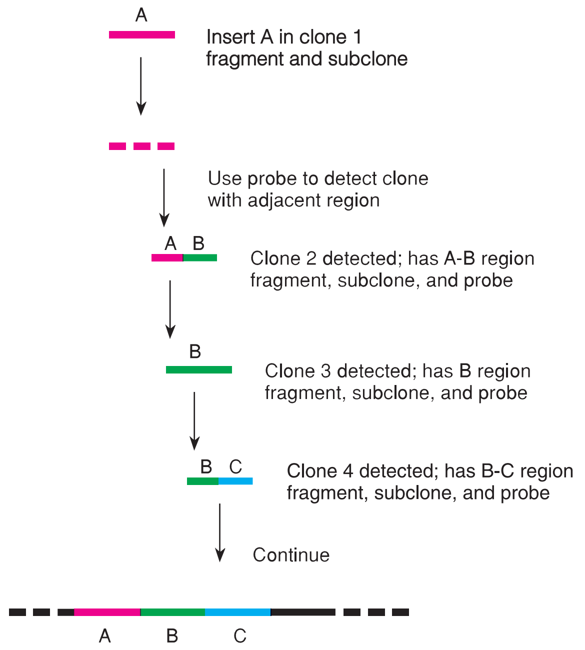

A
Insert A in clone 1 fragment and subclone
Use probe to detect clone with adjacent region
A B
Clone 2 detected; has A-B region fragment, subclone, and probe
B
Clone 3 detected; has B region fragment, subclone, and probe
B C
Clone 4 detected; has B-C region fragment, subclone, and probe
Continue
A
B
C

**Figure 13.38** Chromosome walking technique. This technique allows one to study long chromosomal regions by locating overlapping cloned inserts. We begin with a specific cloned piece of DNA, referred to as insert A. This piece is fragmented to create probes for other clones in a genomic library that contains regions that overlap A (the next region down is referred to as B). The A-B clone is itself then fragmented to create probes to repeat the process, moving down the chromosome.

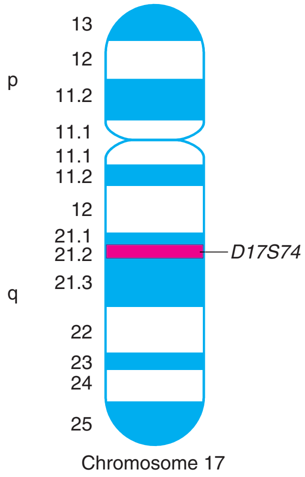

13
12
p
11.2
11.1
11.1
11.2
12
21.1
21.2
D17S74
21.3
q
22
23
24
25
Chromosome 17

**Figure 13.39** A Giemsa-banded chromosome 17, showing the numbering of the regions and the location in region 21q in which marker *D17S74* is located. The terminology of the marker is that of section 74 of chromosome 17. This marker correlated to the position of the *BRCA1* gene.

Two general methods were developed for mapping the human genome, the *standard method,* supported in large part by federal funding, and the **whole-genome shotgun method** used by the Celera Genomics Company. In the standard method, the project is reduced to finding a segment of the genome and locating where it belongs. The segment is then sequenced. By the overlap of sequenced pieces, the whole genome is pieced together. Mapping is done chromosome by chromosome since individual chromosomes can be isolated in large numbers by the methods of flow cytometry, described in chapter 15. In the initial stages of the Human Genome Project, when the primary task was mapping, yeast artificial chromosomes (YACs) were the primary cloning agent. However, as the emphasis of the project shifted

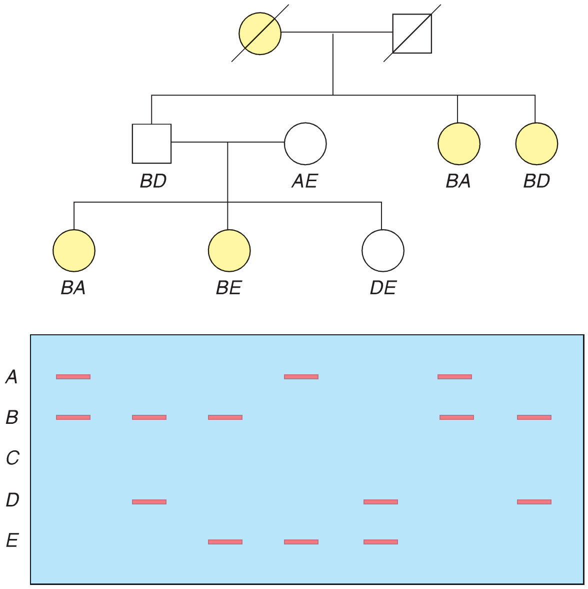

BD
AE
BA
BD
BA
BE
DE
A
B
C
D
E

**Figure 13.40** A pedigree of a family of individuals in which early onset breast cancer is segregating. At the *bottom* of the figure is a gel of the various bands produced, showing the alleles of *D17S74,* marked *A–E* in decreasing size of fragment probed. The individuals in the pedigree are shown directly over their lanes in the gel. The original parents were dead (*diagonal line*) and thus were not typed. The mother, two of her daughters, and two of her granddaughters were diagnosed with breast cancer in ages ranging from twenty-three to forty-five years of age (*yellow*). Note that in every case of breast cancer, the woman has the *B* allele of marker *D17S74.* It is this correlation that localized the breast cancer gene to that region of the chromosome. *D17S74* was the 183rd marker M. King and her colleagues studied; the other markers showed no correlation with breast cancer. (Reprinted with permission from J. M. Hall et al., "Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21," *Science,* 250:1684–89, 1990. Copyright © 1990 American Association for the Advancement of Science.)

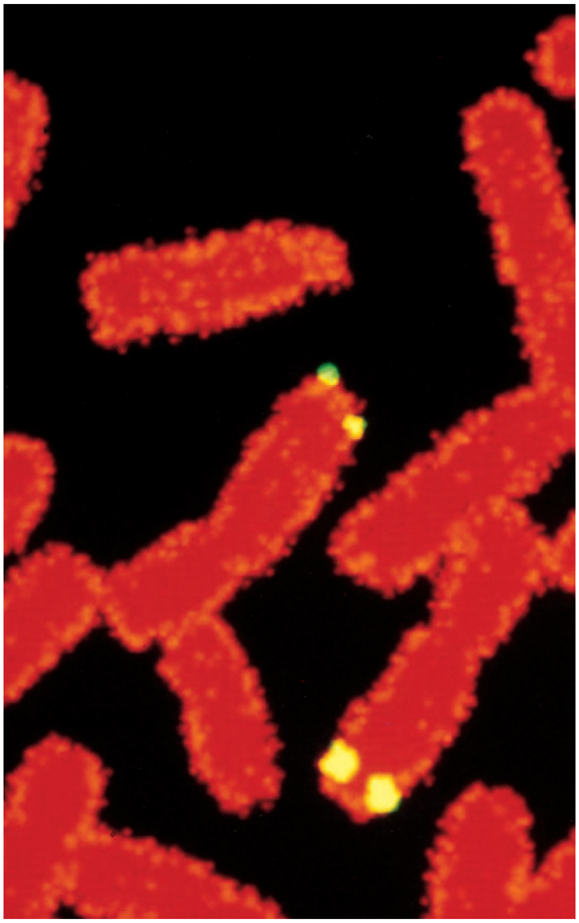


**Figure 13.41** The physical location of a gene or marker can be found by probing chromosomes with a complementary DNA sequence that has a specific fluorescent compound bound to it. When activated, the probe is seen (*bright yellow spots*) in a laser scanning confocal microscope. The chromosomes are counterstained with propidium iodide, which makes them fluoresce *red.* In this case, the probe has located a sequence on human chromosome 11. (© Peter Menzel/Photographed at Yale University Medical School.)

to sequencing, **bacterial artificial chromosomes (BACs)** were used. The bacterial artificial chromosomes are derivatives of the fertility factor (F factor, see chapter 7). They have properties of stability and homogeneity that make them more compatible with automated sequence techniques.

To begin sequencing, each individual chromosome is broken up into overlapping segments of about 150,000 bp in a BAC library. Each BAC is then digested into smaller pieces that are cloned in cosmids or P1 phages digested into smaller pieces for sequencing.

Before we define the techniques further, we should mention that we are not dealing with just one map of the genome, but several different kinds of maps. Although the ultimate goal was the complete DNA sequence of the genome, yielding the exact location of every gene, we needed to go through several stages to get there—remember, we are trying to keep track of 3.3 billion bases. We are familiar with the genetic linkage map of chromosomes described in chapter 6. These maps are called **classical linkage maps;** they define distances in recombination frequencies. A **modern linkage map** is one that uses RFLP markers along its length instead of genes. There is also a **physical map,** in which distances are in physical units of base pairs. These maps can be of microsatellite markers or of **sequence-tagged sites (STSs).** Sequence-tagged sites are DNA lengths of 100-500 base pairs that are unique in the genome. They are created by polymerase chain reaction amplification of primers obtained by sequencing segments of the genome. The primers are then tested to be sure the sequence is unique. About 50% of attempts yield sequence-tagged sites.

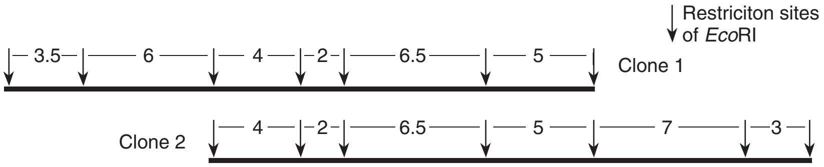
The physical map can also be marked off in differences among individuals that amount to changes in single base pairs. These differences are called **single-nucleotide polymorphisms (SNPs**—pronounced "snips"). These are located about every one thousand bases along the human genome. These single-nucleotide polymorphisms are expected to be especially useful in keeping track of differences among individuals in genes responsible for diseases.

RFLPs, microsatellite markers, STSs, and SNPs allow us to keep track of BACs and cloned pieces in cosmids and P1 phages. However, as we locate various DNA pieces, we will be building up continuous regions of a chromosome by overlapping these pieces. These overlapping, contiguous clones are referred to as **contigs.** This process is repeated chromosome by chromosome. In other words, we are creating a library of overlapping clones that cover the complete length of each chromosome. In essence, we are putting together a linear jigsaw puzzle. Contigs are created by comparing the segments that clones have in common, if any (fig. 13.42). From shared segments, we can infer which parts of the clones overlap. Through this process, contigs of parts of the chromosome can be built up (fig. 13.43). Later, contigs comprising part of a chromosome can be ordered by taking an end clone of a completed contig and using it as a probe to begin chromosome walking to find an end clone of a nearby contig (fig. 13.44).

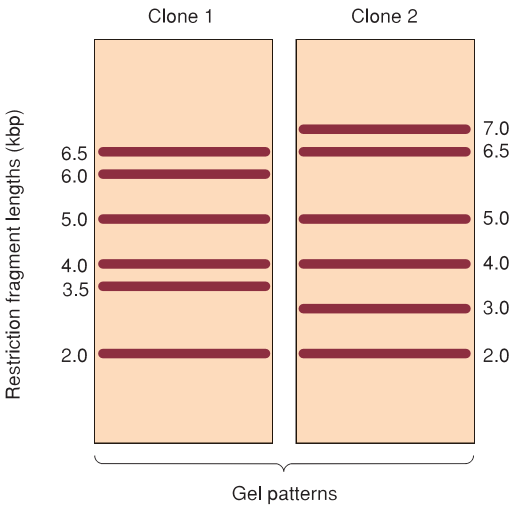
For example, let's begin with a BAC from chromosome 7 of 150,000 base pairs long with three sequence-tagged sites located along its length. We can determine neighboring BACs by shared sequence-tagged sites. The BAC is then digested and cloned into cosmids. The

**Restriction-Fragment Fingerprints**

(a) **Clone 1 overlapping clone 2**


Restriciton sites of *Eco*RI
3.5
6
4
2
6.5
5
Clone 1
Clone 2
4
2
6.5
5
7
3

(b) **Fingerprints of clones 1 and 2**


Clone 1
Clone 2
Restriction fragment lengths (kbp)
6.5
6.0
5.0
4.0
3.5
2.0
7.0
6.5
5.0
4.0
3.0
2.0
Gel patterns

(c) **Regions of overlap and nonoverlap inferred from fingerprint date in (*b*). Fragments are arbitrarily ordered, from largest to smallest, within each region.**

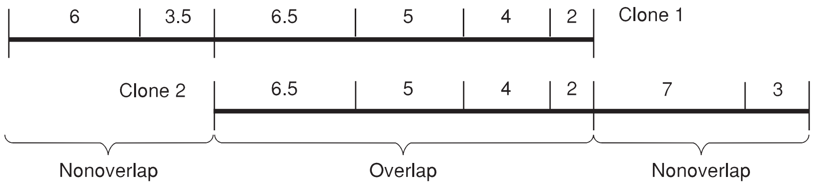

6
3.5
6.5
5
4
2
Clone 1
Clone 2
6.5
5
4
2
7
3
Nonoverlap
Overlap
Nonoverlap

**Figure 13.42** To create contigs, researchers must find overlapping clones and determine their region of overlap. In part (*a*), we have two overlapping pieces of DNA, found by chromosome walking. The pieces are digested with *Eco*RI and electrophoresed, producing the blots in part (*b*). From these gels, we see that fragments of 2.0, 4.0, 5.0, and 6.5 kb pairs are in common, indicating that they are in the region of overlap in both clones. We have thus isolated the overlap region and the unique end regions of both clones (compare *c* with *a*). Restriction maps can then be made of each segment, ordering the pieces. (Reprinted courtesy of *Los Alamos Science,* Volume Number 20, a publication of Los Alamos National Library, Los Alamos, NM.)

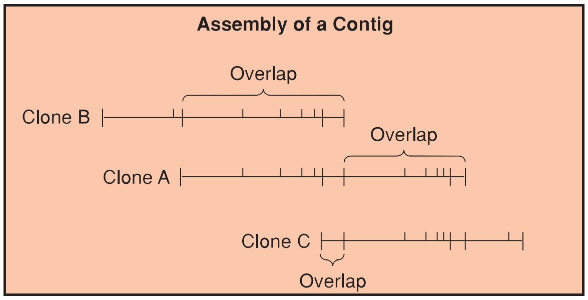

Assembly of a Contig
Overlap
Clone B
Overlap
Clone A
Clone C
Overlap

**Figure 13.43** A contig is further built up by assembling pairwise overlapping clones into longer sequences. Here we see that clone A overlaps clone B to the *left* and clone C to the *right.* In this case, there is one fragment common to all three clones. By comparing clones in this manner, we can march down the chromosome, creating a larger and larger contig. (Reprinted courtesy of *Los Alamos Science,* Volume Number 20, a publication of Los Alamos National Library, Los Alamos, NM.)

overlap of cosmids can be determined by sequence-tagged sites, RFLPs, SNPs, or microsatellites in common. The cosmids are then digested and sequenced. From the sequences we work back, finding overlap and thereby constructing a contig of that BAC. The same process is carried out on neighboring BACs, extending the contig eventually to cover the entire chromosome.

At the initiation of the Human Genome Project, various goals were set. A modern linkage map of microsatellite markers of the human genome was targeted to be complete when markers were spaced about 0.7 centimorgans (about 700,000 base pairs) apart. That goal was reached in 1996 with 2,335 microsatellite markers located on the genome. The physical map of sequence-tagged sites would be considered complete with markers every 100,000 bases, the equivalent of 30,000 sequence-tagged sites in the genome cloned in BACs. That goal was reached in 1997. The sequence of the complete genome was targeted for 2001 and announced in 2000.

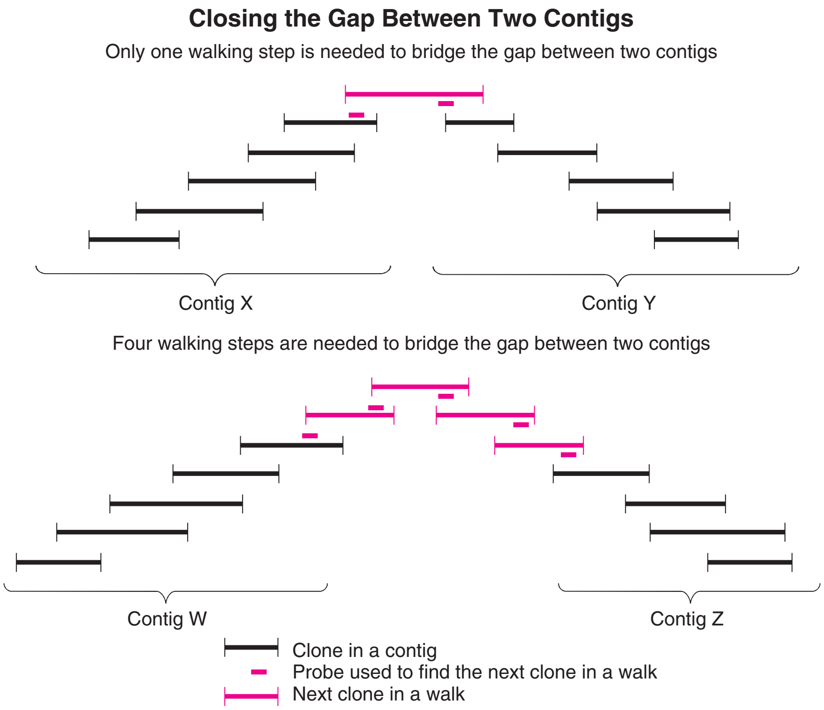

Closing the Gap Between Two Contigs
Only one walking step is needed to bridge the gap between two contigs
Contig X
Contig Y
Four walking steps are needed to bridge the gap between two contigs
Contig W
Contig Z
Clone in a contig
Probe used to find the next clone in a walk
Next clone in a walk

**Figure 13.44** When contigs of large parts of a chromosome are built up, they need to be connected. We can do this directly if there is an overlap at the end of one contig and the beginning of the next. Barring that, we must do chromosome walking to find clones that bridge the gap between two contigs. At the *top* of the figure, in typical chromosome walking technique, the DNA of an end clone is fragmented and used to probe for an overlap. In this case, one clone is found that overlaps two contigs and thus joins them into one long contig. In the *bottom* portion of the figure, the walk requires finding four overlapping clones that bridge the gap between the two contigs. In both cases, the process is successful, joining two contigs into one longer one. (Reprinted courtesy of *Los Alamos Science,* Volume Number 20, a publication of Los Alamos National Library, Los Alamos, NM.)

One of the reasons that goals were optimistic is that methods of mass production have been developed as the project has moved along. These methods include the automation of sequencing and cloning and the development of some new technology. For example, scientists at Affymetrix, Inc., have developed the equivalent of a DNA probe computer chip. Thousands of known DNA sequences are synthesized on a glass substrate. The DNA to be probed is introduced to this chip, where hybridization will take place. Using fluorescent technology, successful probing can be determined using a laser confocal scanning system (fig. 13.45). These chips allow extremely rapid analysis of DNA sequences. Several other manufacturers have developed similar technologies.

As mentioned at the beginning of this chapter, J. C. Venter of Celera Genomics was a co-announcer of the completion of the sequencing of the entire human genome. Venter and his colleagues used a whole-genome shotgun method in which the entire human genome was broken into small segments, cloned, and sequenced. The Celera group will then piece together the genome with a massive computing effort. Previously, it had been thought that this method could not work on a genome as large as the human genome. Venter and his colleagues, however, had sequenced the *Drosophila* genome (180 million base pairs) by March of 2000 by this method. Venter and his colleagues had also sequenced the first true organism, the bacterium *Haemophilus influenzae* (1.8 million base pairs) in July of 1995. Since that time, the yeast *Saccharomyces cerevisiae* (12 million base pairs) was sequenced in 1996, and a significant genetic model organism, the nematode worm, *Caenorhabditis elegans* (97 million base pairs; see chapter 16), was sequenced in 1998. Since 1995, numerous other bacteria and eukaryotes have had their genomes sequenced.

### *Bioinformatics and Proteomics*

These incredible accomplishments in genomics have given rise to two newly named sciences, **bioinformatics** and **proteomics.** Bioinformatics is the science of mining the data from the DNA sequences obtained from sequencing. Mining refers to the storage, retrieval, and analysis of the data. Proteomics is the study of the **proteome,** from *prote*ins of the gen*ome,* and refers to the study of the complete set of proteins from a particular

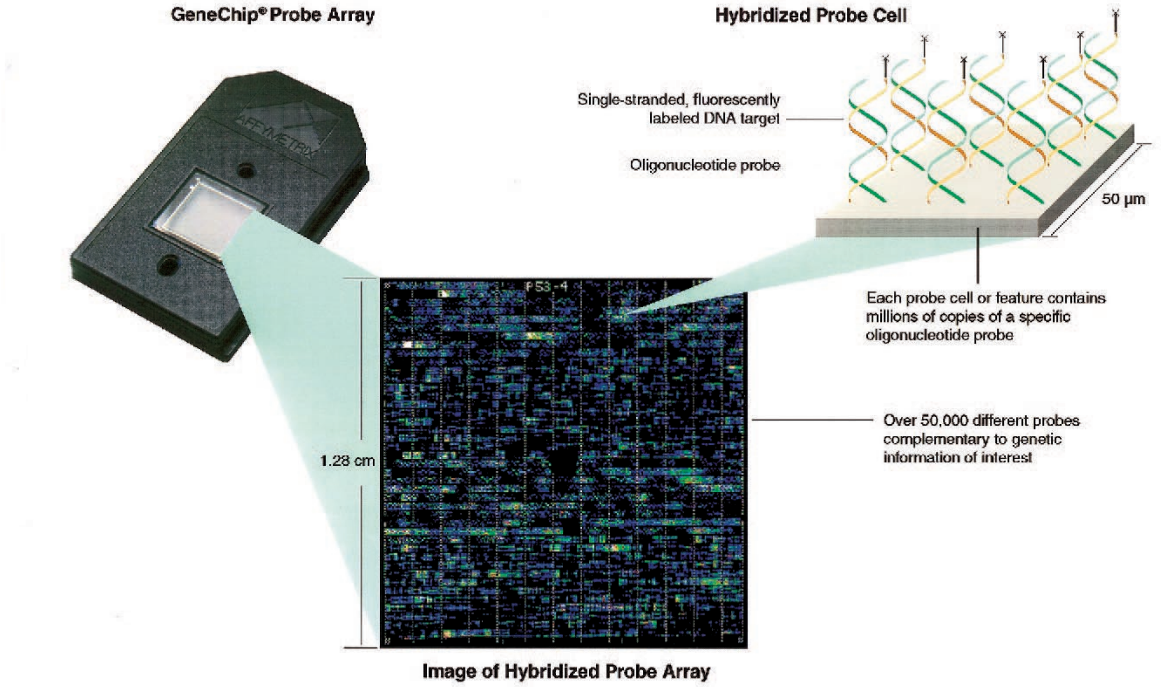

GeneChip® Probe Array
Hybridized Probe Cell
Single-stranded, fluorescently labeled DNA target
Oligonucleotide probe
50 µm
Each probe cell or feature contains millions of copies of a specific oligonucleotide probe
Over 50,000 different probes complementary to genetic information of interest
1.28 cm
P53-4
Image of Hybridized Probe Array

**Figure 13.45** The Gene Chip® DNA probe array is a glass wafer containing from sixty-five thousand to one million or more different DNA sequences. The chips are created by photolithographic techniques, similar to those used in computer chip manufacture. The DNA being probed must have fluorescent molecules attached to permit rapid screening. (Courtesy of Affymetrix, Inc.)

genome. It is the protein analogue to genomics. It is estimated that there are from 50,000 to 2 million different proteins in biological systems, although the number of distinct shapes—motifs—may only be about five thousand. The role of proteomics is to characterize all proteins and determine their structures and shapes. As classical genetics worked from phenotype to genotype, modern molecular genetics is working from genotype (genomics) to phenotype (proteomics).

### Ethics

In addition to the expected scientific and medical information that we will gain from sequencing of the human genome are ethical problems that the project will create. We will shortly have the ability to test people for various genes that we cannot test for now, such as genes for latent diseases like coronary artery disease and cancer. Can insurance companies then demand to test individuals for a whole battery of genes and decide afterward whether that person is insurable or what that person's insurance rates should be? Will many persons find themselves uninsurable because they have genes that might predispose them to cancer? Will individuals find themselves unemployable because of similar problems? What should doctors do about diagnosing a genetic disease (such as Huntington disease) that has no cure? Should they tell the patient? Another ethical issue is the extent to which genetic intervention should be used to change the course of a person's life. With the knowledge of the sequence and location of our genes, and the technology to transfer genes into people, will transgenic people become the norm (see box 13.2)? Should we not only cure diseases this way but tailor a person to some ideal? Will genetic intervention into our basic genetic blueprint be routine? To address these questions, an ethics panel has been set up as part of the Human Genome Project.

## PRACTICAL BENEFITS FROM GENE CLONING

Throughout this chapter, we have mentioned applications of genetic engineering. Here we summarize some of the accomplishments and future directions in the medical, agricultural, and industrial arenas.

### Medicine

In medicine, genetic engineering has had remarkable successes in some areas. On the one hand, basic knowledge about how genes work (and don't work) has advanced tremendously. On the other hand, recombinant DNA methodology has made available large quantities of substances previously in short supply. These include insulin, interferon (an antiviral agent), growth hormone, growth factors, blood-clotting factors, and vaccines for diseases such as hepatitis B, herpes, and rabies. Advances in AIDS and cancer research are discussed in chapter 16. Genetic engineering is making it possible to manufacture antibodies to diagnose and treat diseases. The sequencing of the human genome will further aid medicine by identifying the genes for various diseases, a first step in discovering cures. So far, several genes of great importance have been located, cloned, and sequenced. We also pointed out the use of restriction fragment length polymorphisms and the polymerase chain reaction as techniques of tremendous power in identifying individuals for forensic purposes.

On another front, transgenic mice and cloned sheep have shown that genetic engineering can be applied to higher organisms (fig. 13.46). The use of this technology to treat human diseases, however, is only just beginning. In July 1990, the National Institutes of Health approved gene therapy treatments on people: A child was infused with cells to replace a gene for the enzyme adenosine deaminase, an enzyme whose absence results in a dysfunctional immune system. Although the latter treatment was successful, it had been augmented by other treatments, rendering the conclusions equivocal. Mice and dogs have had hemophilia B corrected by infusion of a genetically modified adenovirus. AIDS, hemophilia, cystic fibrosis, and diabetes are other diseases that should be amenable to gene therapy in the near future.

**Figure 13.46** The sow shown is transgenic, producing large quantities of human protein C in her milk. The protein controls blood clotting and is normally found only in trace quantities in human blood. (Courtesy of William H. Velander, Virginia Tech.)

## Agriculture

Currently, in the United States, approximately one quarter of farmland is planted with crops that are genetically modified. Most are resistant to certain insect pests because they contain genes from *Bacillus thuringiensis* (often referred to as Bt). These genes are for insecticidal proteins called δ endotoxins. For example, the proteins Cry1A and Cry1C from *Bacillus thuringiensis* protect the plants against larval forms of lepidopterans such as the European corn borer. Cry3A protects against coleopterans such as the Colorado potato beetle. In excess of fifty genetically altered crop plants have been approved for planting, including those protected against insect pests, frost, and premature ripening. Rice is being modified so that its vitamin A potential is maintained even after the husks are removed, a procedure done to allow for storage since the husks become rancid. That change alone will improve the health of millions of people throughout the world. Box 13.1 discussed some of the ethical concerns surrounding genetically modified crop plants.

## Industry

Industrial applications of biotechnology include engineering bacteria to break down toxic wastes, modifying yeast to use cellulose to produce glucose and alcohol for fuel, using algae in mariculture (the cultivation of marine organisms in their natural environments) to produce both food and other useful substances, and developing better food processing methods and waste conversion. As an example, baker's yeast (*Saccharomyces cerevisiae*) has been modified with a plasmid that contains two cellulase genes, an endoglucanase and an exoglucanase, that convert cellulose to glucose. The yeast can then convert glucose to ethyl alcohol. These yeasts are now capable of digesting wood (cellulose) and converting it directly to alcohol. The potential exists to harvest the alcohol the yeast produces as a fuel to replace fossil fuels that are in dwindling supply and are polluting the planet.

As you can see, there is no one direction that biotechnology is taking. Many advances are being made that will probably affect every person's life in a beneficial way. Cautious optimism is certainly in order.

# SUMMARY

**STUDY OBJECTIVE 1:** To look at the techniques of gene cloning 359–377

Recombinant DNA techniques revolve around the cloning of foreign DNA in a plasmid or phage. Cloned DNA can be amplified, expressed, and sequenced. Gene cloning techniques came about with the discovery of restriction endonucleases. Type II restriction endonucleases cleave DNA at palindromic regions, which have twofold symmetry.

Recombinant vectors can be constructed several different ways. Foreign and vector DNA can be made compatible by treating each with the same restriction endonuclease—each will then have the same sticky ends. If that does not work, T4 DNA ligase can join blunt ends. In a variation of this method, linkers containing restriction sites are added to vector and foreign DNA. These linkers are then treated with a restriction endonuclease that gives the DNA sticky ends.

DNA to be cloned can be synthesized from an RNA template (cDNA) or isolated by various techniques. If messenger RNA is available, it can be converted into a clonable complementary DNA with the enzyme reverse transcriptase. If DNA is to be isolated directly, it must be identified among all the other DNA fragments created. Locating a desirable piece of DNA is done with probes, complementary nucleic acids labeled with radioactivity or chemiluminescence. Southern blotting, a transfer technique, is used first, followed by DNA-DNA or DNA-RNA hybridization and autoradiography. If the DNA is cloned first, as in the creation of a genomic library, probes can be created or expression of the cloned gene can be determined.

Eukaryotic vectors have been developed, including yeast plasmids, tumor virus vehicles in animals, and crown gall tumor plasmids in plants. Eukaryotes can be transfected by foreign DNA and express it in transgenic organisms. DNA can be injected, shot in on projectiles, electroporated, or introduced by viruses, plasmids, or liposomes. Knockout mice, lacking a specific gene, can be created.

**STUDY OBJECTIVE 2:** To examine the techniques of creating restriction maps 377–383

Restriction digests can be separated by electrophoresis, then used to construct a restriction map. This is a map of the DNA showing the location of restriction enzyme recognition sites. The genetic maps, generated by mating analysis, can then be superimposed on the restriction maps, locating regions of interest on the physical map. Restriction fragment length polymorphisms (RFLPs) provide a tool for locating genes through linkage analysis and are also valuable in forensic science. The polymerase chain reaction (PCR) is a technique used to rapidly amplify particular segments of DNA.

**STUDY OBJECTIVE 3:** To study the methods of DNA sequencing 383–390

DNA is usually sequenced by one of two methods. The dideoxy method developed by Sanger and his colleagues requires the synthesis of DNA in the presence of chain-terminating (dideoxy) nucleotides. Electrophoresis followed by autoradiography allows the sequence of nucleotides synthesized to be determined directly. Fluorescent labeling allows computerized sequence determinations. The phage φX174 was sequenced in its entirety through the forerunner of this technique, the plus-and-minus method. Gilbert and Maxam's chemical method also is used widely.

**STUDY OBJECTIVE 4:** To look at the goals and methods of the Human Genome Project 390–397

The Human Genome Project is a massive, international effort to map and sequence all 3.3 billion bases of the human genome. Initial success was announced in the spring of 2000. Modern linkage maps are being created of restriction sites, microsatellite markers, sequence-tagged sites, and single-nucleotide polymorphisms. They are being coordinated with physical maps created with overlapping contiguous clones of chromosomes. These techniques currently allow us to find genes of interest. The project also includes the sequencing of the genomes of other relevant organisms.

**STUDY OBJECTIVE 5:** To look at the practical benefits and human issues of genetic engineering 397–398

Genetic engineering is moving forward on a number of fronts. Medical, agricultural, and industrial applications are becoming widespread.

## SOLVED PROBLEMS

**PROBLEM 1:** A piece of eukaryotic DNA is obtained by using a restriction endonuclease that leaves blunt ends (*Hae*III). How could we get this piece of DNA into a *Bam*HI site in plasmid pBR322, and how would we know when the foreign DNA has been cloned?

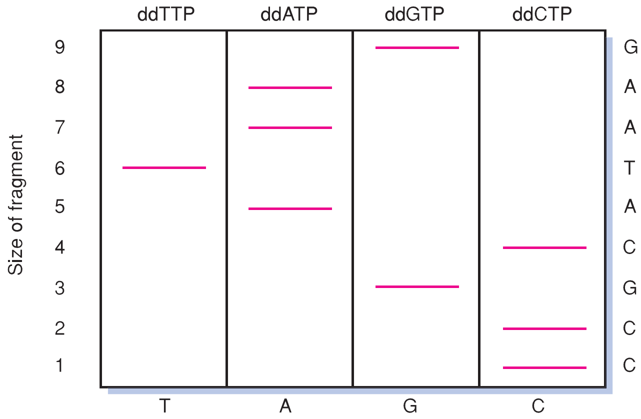
***Answer:*** Since the two pieces of DNA (the eukaryotic piece and the plasmid) have different ends, they must be made compatible before cloning. The simplest way would be to attach blunt-ended linkers to the foreign DNA with phage T4 DNA ligase (see fig. 13.9). The linkers, of course, would have a *Bam*HI site within. After the linkers are attached to the foreign DNA, it would be treated with the *Bam*HI restriction enzyme, giving the foreign DNA *Bam*HI ends. The plasmid is then also treated with the restriction enzyme and the two (the foreign DNA and the cut plasmid) are now mixed together in the presence of *E. coli* DNA ligase, which seals up the plasmids, with or without cloned inserts (see fig. 13.6). Since they have compatible ends, some of the time, a piece of foreign DNA is inserted into a plasmid. The plasmids are then taken up by *E. coli* cells that are grown overnight in an incubator. The bacterial colonies are then replica-plated on media with the antibiotics ampicillin or tetracycline. Colonies that are resistant to ampicillin but sensitive to tetracycline are assumed to be bacteria containing plasmids with cloned inserts (see fig. 13.8).

**PROBLEM 2:** How does a reporter system work?

***Answer:*** A reporter system is a genetically manipulated system that displays a particular phenotype or reaction when a desired event has taken place. In this chapter, we discussed the firefly luciferase reporter system in which the desired result (transcription of a particular promoter) causes a transgenic tobacco plant to glow. Let us say that we are studying the control of transcription of a particular eukaryotic gene. We could attach the promoter of that gene to the firefly luciferase gene in a Ti plasmid by cloning techniques. The plasmid could then transfect tobacco plants, and we could continue our experiment to determine whether the promoter under study is active under various conditions. We would know whether it was active by watering the plants with luciferin. If the plant glows, then the luciferase gene product is present, which means that the promoter under question is active. In other words, the glowing of the plant "reports" the action of the promoter under question; the promoter is active because it allowed the transcription of the luciferase gene. We also discussed the green fluorescent protein reporter system.

**PROBLEM 3:** A piece of DNA has the sequence 3′-GGCGTATTC-5′. It is sequenced using the dideoxy method. How many bands are found on the ladder gel? How many bands and of what size are found for each reaction mixture?

***Answer:*** Since the piece of DNA is nine bases long, the total number of bands in all four lanes of a sequencing gel add up to nine (see fig. 13.33). By each reaction mixture, we mean the four reaction mixtures each with one of the dideoxynucleotides. In the reaction mixture with ddTTP, chain termination occurs at the adenine in the piece of DNA; that is, a DNA segment was synthesized that is six bases long (see the following figure). In the reaction mixture with ddATP, chain termination occurs opposite each of the thymines, producing DNA segments of five, seven, and eight nucleotides. In the reaction mixture with ddGTP, chain termination occurs opposite the cytosines in positions three and nine. And, in the reaction mixture with ddCTP, chain termination occurs after synthesis of segments one, two, and four bases long. Note that the gel gives us the sequence of the complement strand of the original piece of single-stranded DNA.



**PROBLEM 4:** A linear DNA molecule 1,000 base pairs long is digested with the following restriction enzymes, producing the following results:

| | |
|---|---|
| *Eco*RI | 400 bp, 600 bp |
| *Bgl* II | 250 bp, 750 bp |
| *Eco*RI + *Bgl* II | 250, 350, 400 bp |

Determine the restriction map.

***Answer:*** Each enzyme alone produces two fragments, so the molecule has one site for each enzyme. Since we get different-sized fragments with each enzyme, the sites must be located asymmetrically along the DNA. Draw these sites:

*Eco*RI *Bgl*II
400 ↓ 600 and 250 ↓ 750

The *Eco*RI fragment that lacks a *Bgl*II site should appear in the double digest. If *Bgl*II cuts within the 400 base-pair fragment, we would expect to see 150, 250, and 600 base-pair fragments. We don't see this, so the *Bgl*II site is not within the 400 base-pair *Eco*RI fragment. Thus, the map looks like this:

*Eco*RI *Bgl*II
400 ↓ 350 ↓ 250

## EXERCISES AND PROBLEMS*

### GENOMIC TOOLS

1. What specific properties of type II endonucleases make them useful in gene cloning?
2. The following is a double helix of DNA. What, if any, are potential restriction enzyme recognition sequences?

   5′-TAGAATTCGACGGATCCGGGGCATGCAGATCA-3′
   3′-ATCTTAAGCTGCCTAGGCCCCGTACGTCTAGT-5′

3. Assuming a random arrangement of nucleotides on a piece of DNA, what is the probability that a restriction endonuclease whose recognition site consists of four bases (a four-cutter) will cut the DNA? What is the probability for a six-cutter? an eight-cutter?
4. Under what circumstances is a restriction endonuclease unsuitable for cloning a piece of foreign DNA?
5. What methods exist to create sticky ends or create ends for joining two incompatible pieces of DNA? When is each method favored?
6. Diagram a possible heteroduplex between two phage λ vectors, one with and one without a cloned insert, created by DNA-DNA hybridization.
7. What are the differences among plasmid, cosmid, expression vector, and YAC? Under what circumstances is each useful?
8. What are the steps by which messenger RNA can be converted into cDNA? How would we obtain radioactive cDNA? radioactive messenger RNA?
9. What is chromosome walking? When is it used?
10. How would we isolate a human alanine transfer RNA gene for cloning? How would we locate a clone with a human alanine transfer RNA gene in a genomic library?
11. What are the differences among Southern, western, northern, and dot blotting?

---

*Answers selected exercises and problems are on page A-14.

**12.** How would you develop a probe for a gene whose messenger RNA could not be isolated? How could an expression vector be used to isolate a cloned gene?

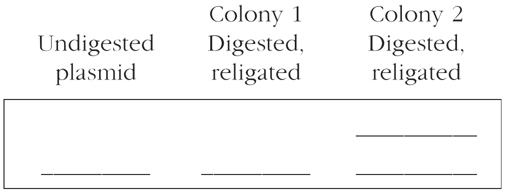
**13.** How are *E. coli* plasmids manipulated to survive in yeast? How can virus genomes, such as SV40 and phage λ, survive as functioning vectors when parts of their genomes are replaced by cloned DNA?

**14.** What methods are used to get foreign DNA into eukaryotic cells? What is transfection? What is a transgenic mouse?

**15.** Exonuclease III is an enzyme that sequentially removes bases from the 3′ end of double-stranded DNA. The following two molecules, each 100 bp long, are digested with exonuclease III. Molecule 1 is completely digested; molecule 2 is only partially digested. Explain these results.

Molecule 1: CGTTCAG...
GCAAGTC...
Molecule 2: AAAAAAAAAA...
TTTTTTTTTT...

**16.** A plasmid that contains an *Eco*RI site within a gene for ampicillin resistance is cut with *Eco*RI, and then religated. This plasmid is used to transform *E. coli* cells, and the plasmid is reisolated from the ampicillin-resistant colonies. The reisolated plasmids from two different colonies are electrophoresed, and the results appear in the following figure.



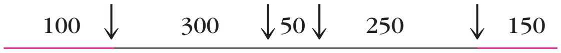
How do you account for the two bands in colony 2?

**17.** Most human genes contain one or more introns. Since bacteria cannot excise introns from nuclear messenger RNA (snRNPs are needed), how can bacteria be used to make large quantities of a human protein?

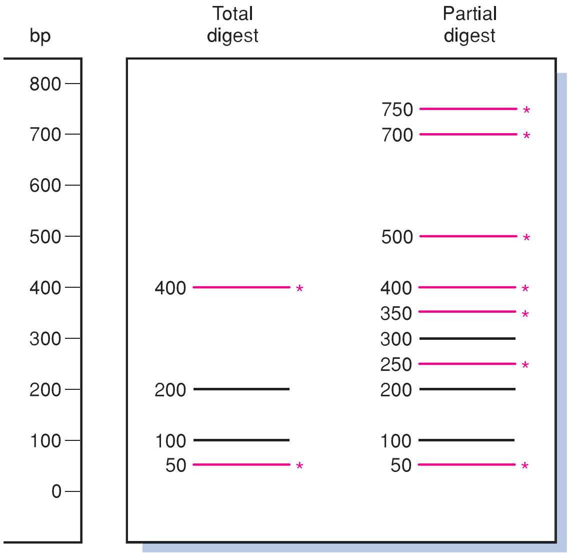
## RESTRICTION MAPPING

**18.** How are DNA fingerprints useful in forensic cases? Could they be used in paternity exclusion?

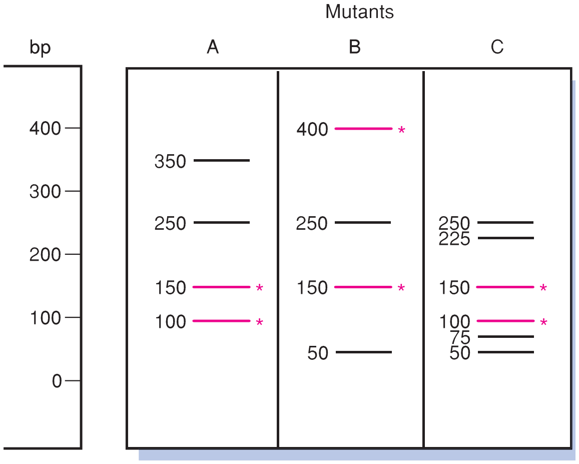
**19.** The following segment of DNA is cut four times by the restriction endonuclease *Eco*RI at the places shown. Diagram the gel banding that would result from electrophoresis of the total and partial digests. Note the end-labeled segments and regions where several segments form bands at the same place on the gel.



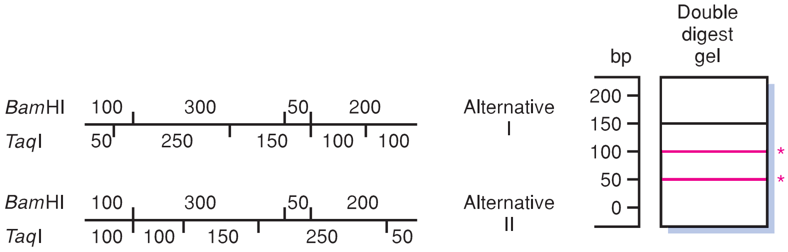
**20.** The following figure shows a gel of a total and partial digest of a DNA segment treated with *Hin*dII. End-labeled segments are noted by asterisks. Draw the restriction map of the original segment.



**21.** Several mutants of the DNA segment shown in problem 19 were isolated. They gave the following gel patterns when the total digests were electrophoresed. Asterisks denote the end-labeled segments. Can you determine the nature of the mutations?



**22.** Restriction maps of a segment of DNA were worked out separately for *Bam*HI and *Taq*I. Two overlays of the maps are possible. The double-digest gel is shown in the following figure (asterisks denote end labels). Which overlay is correct?



**23.** A linear DNA molecule 1,000 bp long gives the following size fragments when treated with these restriction enzymes. Derive a restriction map.

| | |
|---|---|
| *Eco*RI: | 300 bp, 700 bp |
| *Bam*HI: | 150 bp, 200 bp, 250 bp, 400 bp |
| *Eco*RI + *Bam*HI: | 50 bp, 100 bp, 200 bp, 250 bp, 400 bp |

**24.** A linear DNA molecule cut with *Eco*RI yields fragments of 3 kb, 4.2 kb, and 5 kb. What are the possible restriction maps?

**25.** You have double-stranded DNA that you radioactively label at the 5′ ends. Digestion of this molecule with either *Eco*RI or *Bam*HI yields the following fragments. The numbers are in kilobases (kb), and an asterisk indicates the fragments that are labeled.

*Eco*RI: 2.8, 4.6, 6.2*, 7.4, 8.0*

*Bam*HI: 6.0*, 10.0*, 13.0

If unlabeled DNA is digested with both enzymes simultaneously, the following fragments appear: 1.0, 2.0, 2.8, 3.6, 6.0, 6.2, 7.4. What is the restriction map for the two enzymes?

**26.** A 12 kb DNA molecule cut with *Eco*RI yields one 12 kb fragment. When the original molecule is cut with *Bam*HI, three fragments of 2 kb, 4.5 kb, and 5.5 kb are produced. When the fragment from *Eco*RI is treated with *Bam*HI, four fragments of 2 kb, 2.5 kb, 3.0 kb, and 4.5 kb are produced. Draw a restriction map.

**27.** A plasmid 3 kb in length contains a gene for ampicillin resistance and a gene for tetracycline resistance. The plasmid has a single site for each of the following enzymes: *Eco*RI, *Bgl*II, *Hin*dIII, *Pst*I, and *Sal*I. If DNA is cloned into the *Eco*RI site, resistance to either antibiotic is not affected. DNA cloned into the *Bgl*II, *Hin*dIII, or *Sal*I sites abolishes tetracycline resistance, and DNA inserted into the *Pst*I site eliminates ampicillin resistance. If the plasmid is digested completely with enzyme mixes, the following fragments result:

| Mixture | Fragment Size (kb) |
|---|---|
| *Eco*RI + *Pst*I | 0.7, 2.3 |
| *Eco*RI + *Bgl*II | 0.3, 2.7 |
| *Eco*RI + *Hin*dIII | 0.08, 2.92 |
| *Eco*RI + *Sal*I | 0.85, 2.15 |
| *Eco*RI + *Bgl*II + *Pst*I | 0.3, 0.7, 2.00 |

Draw a restriction map of the plasmid, and indicate the locations of the resistance genes and the sites of enzymatic cleavage.

**28.** A gene has the following *Eco*RI restriction map (in kilobases):

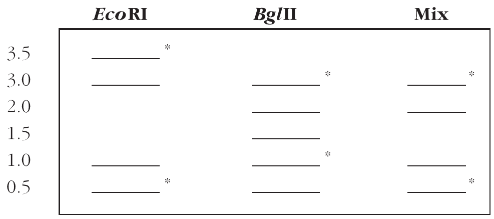
1.0 ↓ 0.7 ↓ 2.0

Draw the gel pattern expected from

**a.** a mutant that has lost the site between the 1.0 and 0.7 kb fragments.

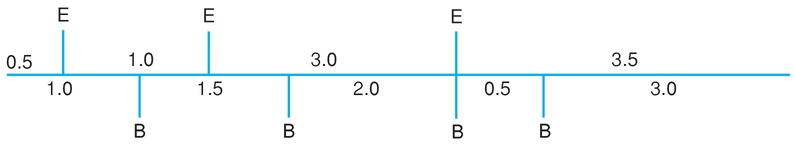
**b.** a mutant that has a new site within the 2.0 kb fragment.

**29.** A DNA fragment 8 kb in size is labeled with $^{32}$P at the 5′ ends. It is then digested with *Eco*RI, *Bgl*II, or a mixture of both enzymes. The size of the fragments and the labeled fragments (*) appear as follows. Sizes are in kilobases.

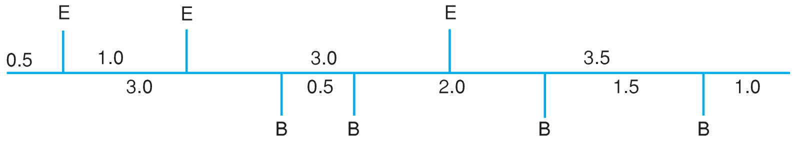


Which of the following two maps is consistent with the results?

1.

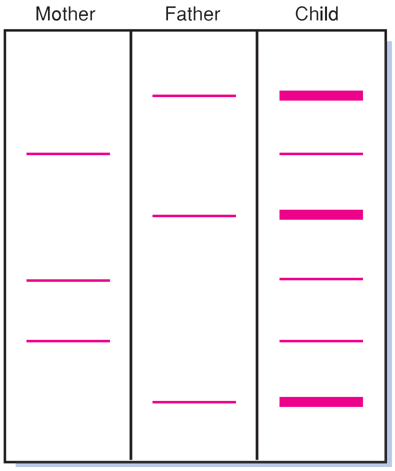

E
E
E
0.5
1.0
3.0
3.5
1.0
1.5
2.0
0.5
3.0
B
B
B
B

2.


E
E
E
0.5
1.0
3.0
3.5
3.0
0.5
2.0
1.5
1.0
B
B
B
B

**30.** You now take an unlabeled molecule from problem 29, digest it with *Hin*dIII, and get two fragments, 5.5 and 2.5 kb in size. If *Hin*dIII does not cut within the 3.5 kb *Eco*RI fragment, what size fragments do you expect in a double digest of *Hin*dIII and *Eco*RI?

**31.** Two normal individuals have a child with Down syndrome. RFLP analysis with a probe from chromosome 21 is performed on all three individuals, and the results of the gels appear as follows. Based on these results, what can you conclude about the origins of the number 21 chromosomes?


Mother
Father
Child

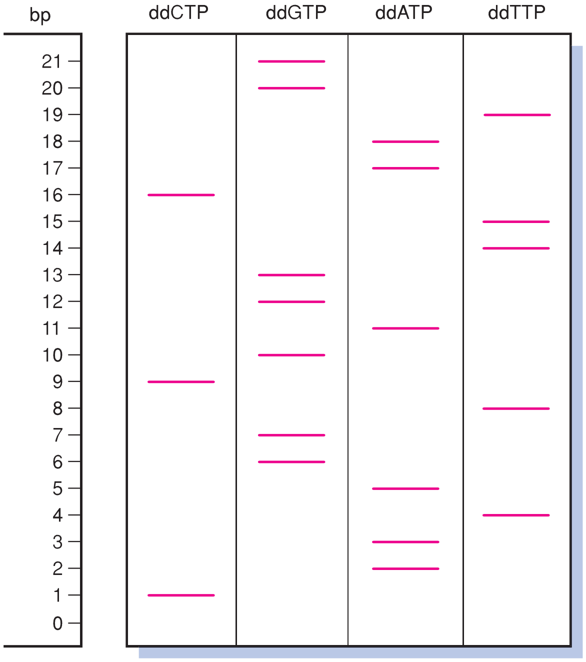
### POLYMERASE CHAIN REACTION

**32.** What is PCR? When is it used?

### DNA SEQUENCING

**33.** What are the steps in the dideoxy method of DNA sequencing? How has the technique been improved with fluorescent dyes?

**34.** The following diagram is of a dideoxy sequencing gel. What is the sequence of the DNA under study?


bp
ddCTP
ddGTP
ddATP
ddTTP
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0

**35.** How can a particular piece of DNA be manipulated to be in the appropriate configuration for dideoxy sequencing?

**36.** Provide, if possible, DNA sequences that can mark the termination of one gene and the initiation of another, given that the genes overlap in one, two, three, four, five, six, or seven bases.

**37.** Draw the expected gel pattern derived from the dideoxy sequencing method for a template strand with the following sequence:

5′-CAGCGAATGCGGAA-3′

**38.** A DNA strand with the sequence 3′-GACTATTCCGAAAC-5′ is sequenced by the dideoxy method. If the reaction mixture contains all four radioactive deoxynucleotide triphosphates plus dideoxythymidine, what size labeled bands do you expect to see on the gel?

### MAPPING AND SEQUENCING THE HUMAN GENOME

**39.** What is hypervariable DNA? a RFLP? a VNTR locus? microsatellite DNA? a sequence-tagged site? (*See also* RESTRICTION MAPPING)

### PRACTICAL BENEFITS FROM GENE CLONING

**40.** Describe some areas of practical benefit from genetic engineering. Why might some people be concerned about its widespread use?

## CRITICAL THINKING QUESTIONS

**1.** In the past, geneticists have used several different methods to splice pieces of DNA that do not have compatible "sticky ends." We mentioned blunt-end ligation and the addition of linkers containing specific restriction sites. Given that nucleotides can be added to the 3′ ends of double-stranded DNA with the enzyme deoxynucleoside terminal transferase, can you see another way to create compatible ends on foreign and vehicle DNA?

**2.** The motion picture *Jurassic Park* was based on the premise that DNA of dinosaurs could be extracted from the blood-meals of mosquitoes preserved in amber and inserted into the genome of a frog, which would then produce living dinosaurs. Is this premise reasonable?

*Suggested Readings for chapter 13 are on page B-11.*

# 14

# GENE EXPRESSION

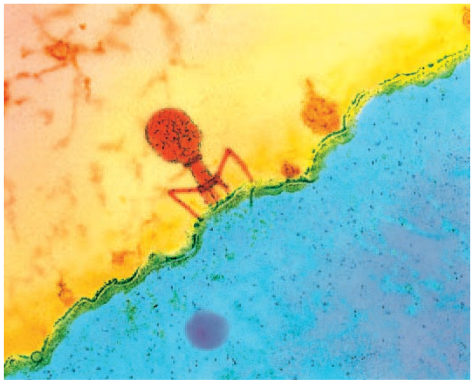
## *Control in Prokaryotes and Phages*

## STUDY OBJECTIVES

1. To study the way in which inducible and repressible operons work 406
2. To examine attenuator control in bacteria 415
3. To analyze the control of the life cycle of phage λ 418
4. To determine the way in which transposable genetic elements transpose and control gene expression in bacteria 425
5. To look at other transcriptional and posttranscriptional mechanisms of control of gene expression in bacteria and phages 430

## STUDY OUTLINE

**The Operon Model 406**
***Lac* Operon (Inducible System)** 406
  Lactose Metabolism 406
  The Regulator Gene 406
  The Operator 408
  Induction of the *Lac* Operon 409
  *Lac* Operon Mutants 409
**Catabolite Repression** 412
***Trp* Operon (Repressible System)** 413
  Tryptophan Synthesis 413
  Operator Control 414
***Trp* Operon (Attenuator-Controlled System)** 415
  Leader Transcript 415
  Leader Peptide Gene 416
  TRAP Control 417
  Redundant Controls 418
**Lytic and Lysogenic Cycles in Phage λ** 418
  Phage λ Operons 418
  Early and Late Transcription 420
  Repressor Transcription 421
  Maintenance of Repression 421
  Lysogenic Versus Lytic Response 423
**Transposable Genetic Elements** 425
  IS Elements 425
  Composite Transposons 427
  Mechanism of Transposition 427
  Phenotypic and Genotypic Effects of Transposition 429
**Other Transcriptional Control Systems** 430
  Transcription Factors 430
  Promotor Efficiency 430
**Translational Control** 430
**Posttranslational Control** 433
  Feedback Inhibition 433
  Protein Degradation 433
*Summary* 434
*Solved Problems* 435
*Exercises and Problems* 436
*Critical Thinking Questions* 438



Artificially colored transmission electron micrograph of a T4 bacteriophage attached to an *Escherichia coli* bacterium. (© Biozentrum, University of Basel/SPL/Photo Researchers, Inc.)

Genes are transcribed into RNA, which, for the most part, is then translated into protein. Control mechanisms are exercised along the way. Without some control of gene expression, an *Escherichia coli* cell, for example, would produce all its proteins in large quantities all the time, and all the cells in a eukaryotic organism would be identical. Although most control mechanisms are negative (preventing something from happening), controls can also be positive (causing some action to occur or enhancing some action). This chapter is devoted to analyzing control processes in prokaryotes and phages; in chapter 16, we examine control processes in eukaryotes.

In the process leading from a sequence of nucleotides in DNA to a protein, control is exerted in many places. In general, control of gene expression can take place at the levels of transcription, translation, or protein functioning. The most efficient place to control gene expression is at the level of transcription.

One of the best-understood mechanisms exerts control of transcription, regulating the production of messenger RNA according to need. *E. coli* messenger RNAs are short-lived in vivo: They degrade enzymatically within about two minutes. A complete turnover (degradation and resynthesis) in the cell's messenger RNA occurs rapidly and continually, and this rapid turnover is a prerequisite for transcriptional control, a central feature of the regulation of prokaryotic gene expression.

## THE OPERON MODEL

Not all of the proteins prokaryotes can produce are needed in all circumstances in the same quantities. For example, some metabolites, such as sugars, which the cell breaks down for energy and as a carbon source, may not always be present in the cell's environment. If a given metabolite is not present, enzymes for its breakdown are not useful, and synthesizing these enzymes is wasteful. If the cell produces enzymes for the degradation of a particular carbon source only when this carbon source is present in the environment, the enzyme system is known as an **inducible system.** Inducible enzymes are synthesized when the environment includes a substrate for those enzymes. The enzymes will then catabolize (break down) the substrate.

On the other hand, the enzymes in many synthetic pathways are in low concentration or absent when an adequate quantity of the end product of the pathway is already available to the cell. That is, if the cell encounters an abundance of the amino acid tryptophan in the environment or if it is overproducing tryptophan, the cell stops the manufacture of tryptophan until a need arises again. A **repressible system** is a system of enzymes whose presence is repressed, stopping the production of the end product when it is no longer needed. Repressible systems are repressed by an excess of the end product of their synthetic (anabolic) pathway.

The best-studied inducible system is the ***lac* operon** in *E. coli.* Since the term *operon* refers to the control mechanism, we will defer a definition until we describe the mechanism.

## *LAC* OPERON (INDUCIBLE SYSTEM)

### *Lactose Metabolism*

Lactose (milk sugar—a disaccharide) is a β-galactoside that *E. coli* can use for energy and as a carbon source after it is broken down into glucose and galactose. The enzyme that performs the breakdown is **β-galactosidase** (fig. 14.1). (The enzyme can additionally convert lactose to allolactose, which, as we will see, is also important.) There are very few molecules of β-galactosidase in a wild-type *E. coli* cell grown in the absence of lactose. Within minutes after adding lactose to the medium, however, this enzyme appears in quantity within the bacterial cell. When the synthesis of β-galactosidase (encoded by the *lacZ,* or *z* gene) is induced, the production of two additional enzymes is also induced: **β-galactoside permease** (encoded by the *lacY,* or *y* gene) and **β-galactoside acetyltransferase** (encoded by the *lacA,* or *a* gene). The permease is involved in transporting lactose into the cell. The transferase is believed to protect the cell from the buildup of toxic products created by β-galactosidase acting on other galactosides. By acetylating galactosides other than lactose, the transferase prevents β-galactosidase from cleaving them.

### *The Regulator Gene*

Not only are the three *lac* genes (*z, y, a*) induced together, but they are adjacent to one another in the *E. coli* chromosome; they are, in fact, transcribed on a single, polycistronic messenger RNA (fig. 14.2). Induction involves the protein product of another gene, called the **regulator gene,** or *i* gene (*lacI*). Although the regulator gene is located adjacent to the three other *lac* genes, it is a totally independent transcriptional entity. The regulator specifies a protein, called a **repressor,** that interferes with the transcription of the genes involved in lactose metabolism.

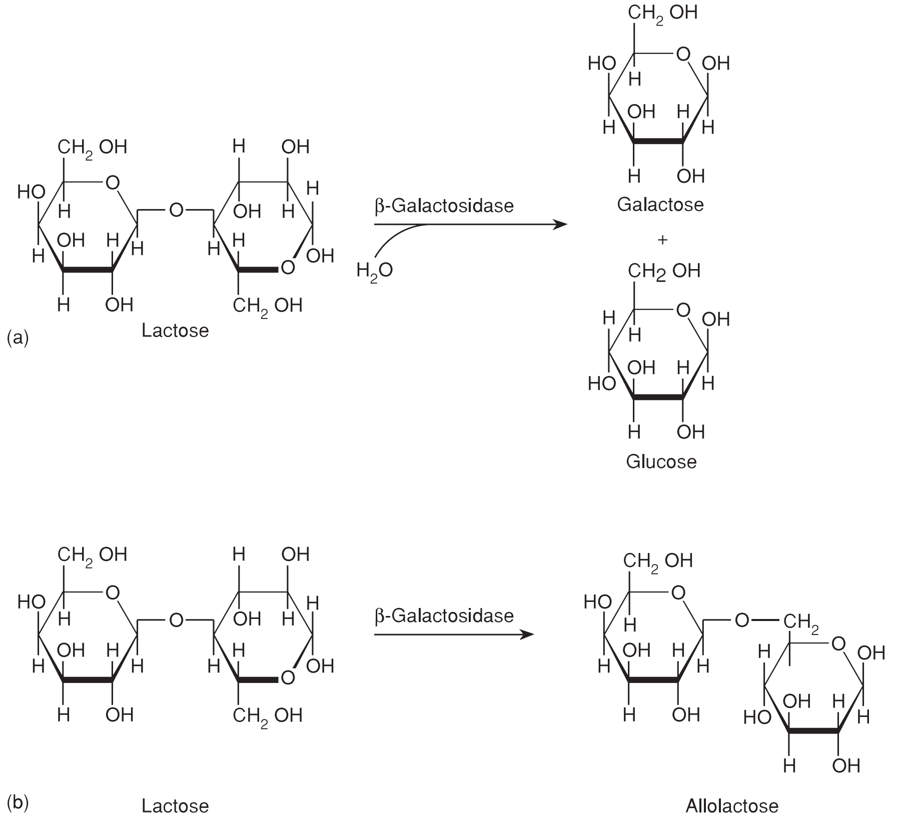

(a)
Lactose
β-Galactosidase
$H_2O$
Galactose
+
Glucose
(b)
Lactose
β-Galactosidase
Allolactose

**Figure 14.1** The enzyme β-galactosidase hydrolytically cleaves lactose into glucose and galactose (*a*). The enzyme can also convert lactose to allolactose (*b*).

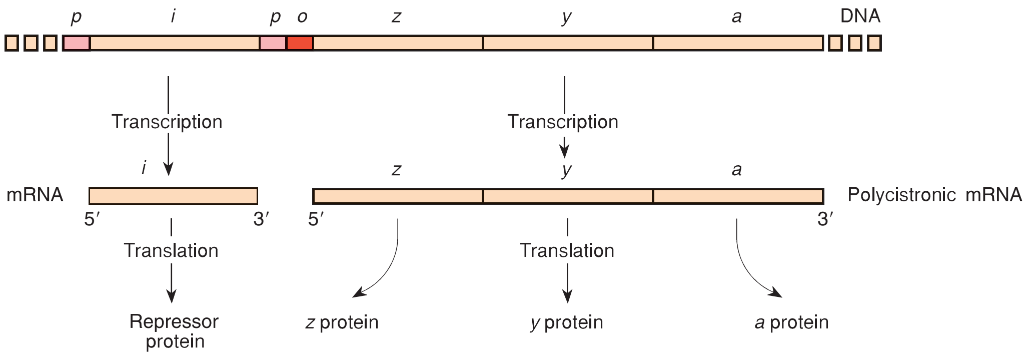

p
i
p
o
z
y
a
DNA
Transcription
Transcription
i
z
y
a
mRNA
Polycistronic mRNA
5′
3′
5′
3′
Translation
Translation
Repressor protein
z protein
y protein
a protein

**Figure 14.2** The *lac* operon is transcribed as a multigenic (polycistronic) mRNA. The *z, y,* and *a* indicate the *lacZ, lacY,* and *lacA* loci. The mRNA transcript is then translated as individual proteins. The *lac* operon regulator gene is denoted as *i*; the *o* stands for operator and the *p* for promoter. Both the operon and the regulator gene have their own promoters. (Source: Data from R. C. Dickson, et al., "Genetic regulation: The *lac* control region," *Science,* 187:27–35, January 10, 1975.)

## *The Operator*

For the repressor protein to exert its influence over transcription, there must be a control element (receptor site) located near the beginning of the β-galactosidase (*lacZ*) gene. This control element is a region referred to as the **operator,** or operator site (fig. 14.2). The operator site is a sequence of DNA that the product of the regulator gene, the repressor, recognizes. When the repressor is bound to the operator, it either interferes with RNA polymerase binding or prevents the RNA polymerase from achieving the open complex (see chapter 10). In either case, transcription of the operon is prevented (fig. 14.3). The repressor is released when it combines with an *inducer,* a derivative of lactose called allolactose (see fig. 14.1).

Note that the promoter not only is recognized by RNA polymerase but also has other controlling elements in the immediate vicinity of the initiation site of transcription. We can now define an **operon** as a sequence of adjacent genes all under the transcriptional control of the same promoter and operator.

The nucleotide sequence of the *lac* operator region is shown in figure 14.4. The operator in figure 14.3 is referred to as the primary operator, $o_1$, centered at +11. Two other operator sequences have been found. One, $o_2$, is centered at +412. The third overlaps the C-terminal end of the *i* gene, is centered at −82, and is referred to as $o_3$. The structure of the repressor and its interaction with the operator sites was worked out recently with X-ray crystallography. The functional repressor is a homotetramer of the protein product of

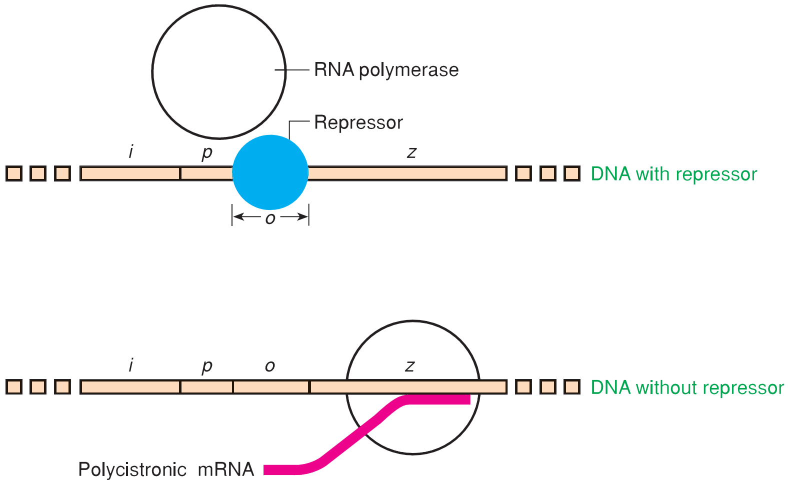

RNA polymerase
Repressor
i
p
z
o
DNA with repressor
i
p
o
z
DNA without repressor
Polycistronic mRNA

**Figure 14.3** The repressor. By binding to the operator, the repressor either prevents RNA polymerase from binding to the promoter and transcribing the *lac* operon as shown, or prevents the polymerase from achieving the open configuration. In either case, transcription of the *lac* operon is prevented. When the repressor is not present, transcription takes place. The functional repressor is a tetramer.

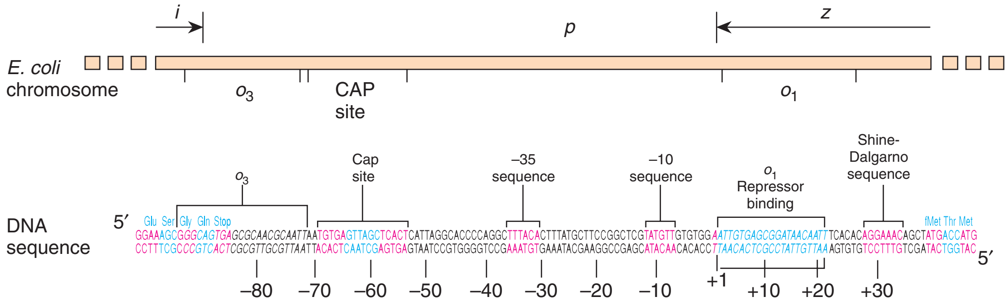

i
p
z
E. coli chromosome
o3
CAP site
o1
DNA sequence
o3
Cap site
−35 sequence
−10 sequence
o1 Repressor binding
Shine-Dalgarno sequence
5′
Glu Ser Gly Gln Stop
GGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATG
CCTTTCGCCCGTCACTCGCGTTGCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGGTCCGAAATGTGAAATACGAAGGCCGAGCATACAACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTTTGTCGATACTGGTAC
fMet Thr Met
5′
−80
−70
−60
−50
−40
−30
−20
−10
+1
+10
+20
+30

**Figure 14.4** The *lac* operon promoter and operator regions. The CAP site is described later. The base sequence corresponds to the diagram above it. The terminal amino acids of the *i* gene are shown, as well as the initial amino acids of the *lacZ* gene. In addition, we picture the Shine-Dalgarno sequence of the DNA, the repressor-binding region (centered at around +10 of the gene), the −10 and −35 sequences of the promoter, and primary ($o_1$) and secondary ($o_3$) operator sites (see text). (Data from R. C. Dickson, et al., "Genetic regulation: The *lac* control region," *Science,* 187:27–35, January 10, 1975.)

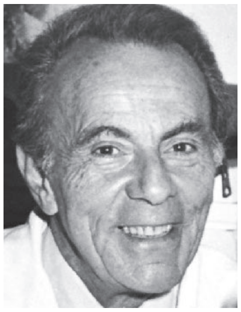
the *i* gene; that is, it is formed from four identical copies of the repressor protein. Since each operator site has twofold symmetry, two repressor monomer proteins bind to each operator site. The monomer is shaped so that it fits into the major groove of the DNA to locate the exact base sequence of the operator; it then binds at that point through electrostatic forces. A tetramer can bind to two of the operator sites at the same time, presumably $o_1$ and $o_3$ or $o_1$ and $o_2$. In the process, the DNA is formed into a loop (fig. 14.5).

## *Induction of the* Lac *Operon*

Under conditions of repression, before the operon can be "turned on" to produce lactose-utilizing enzymes, the repressor will have to be removed from the operator. The repressor is an **allosteric protein;** when it binds with one particular molecule, it changes the shape of the protein, which changes its ability to react with a second particular molecule. Here the first molecule is the inducer allolactose and the second molecule is the operator DNA. When allolactose is bound to the repressor, it causes the repressor to change shape and lose its affinity for operator sequences (fig. 14.5).

With allolactose bound to the repressor, the ability of the repressor to bind to the operator is greatly reduced, by a factor of $10^3$. Since no covalent bonds are involved, the repressor simply dissociates from the operator. After the repressor releases from the operator, RNA polymerase can now begin transcription. The three *lac* operon genes are then transcribed and subsequently translated into their respective proteins.

This system of control is very efficient. The presence of the lactose molecule permits transcription of the genes of the *lac* operon, which act to break down the lactose. After all the lactose is metabolized, the repressor returns to its original shape and can again bind to the operator. The system is "turned off." Using very elegant genetic analysis, details of this system were worked out by François Jacob and Jacques Monod, who subsequently won 1965 Nobel prizes for their efforts.



François Jacob (1920– ).
(Courtesy of Dr. François Jacob.)

Jacques Monod (1910–1976).
(Archives Photographiques, Musée Pasteur.)

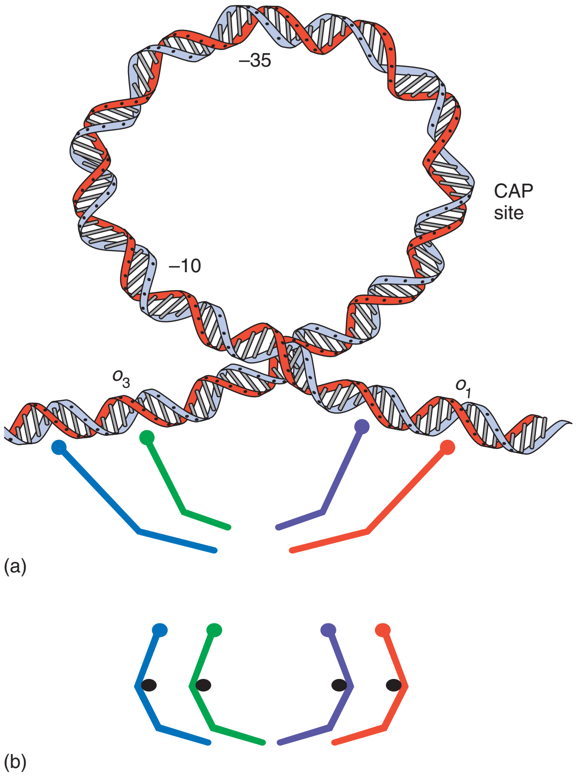


**Figure 14.5** Because the *lac* operator DNA sequences are palindromes, each half can bind one repressor subunit. (*a*) The tetrameric repressor binds to $o_1$ and $o_3$, causing the DNA in between to form a loop. Each of the subunits is shown in a different *color.* The round portion of the subunit in touch with the DNA is the N-terminal end of the repressor subunit; the C-terminal ends form tails that bind the subunits together. Also indicated are the CAP site and the −10 and −35 sequences. (*b*) When each of the subunits binds an allolactose molecule (*black circles*), the shape of the middle portion of the subunit changes, causing the subunit to fall free of the operators.

## Lac *Operon Mutants*

### *Merozygote Formation*

Discovery and verification of the *lac* operon system came about through the use of mutants and partial diploids of the *lac* operon well before DNA sequencing techniques had been developed. The structural (enzyme-specifying) genes of the *lac* operon, *z, y,* and *a,* all have known mutant forms in which the particular enzyme does not perform its function. These mutant forms are designated $z^-$, $y^-$, and $a^-$. The alleles for normal forms of the enzymes are $z^+$, $y^+$, and $a^+$.

Partial diploids in *E. coli* can be created through sexduction (chapter 7) because some strains of *E. coli* have the *lac* operon incorporated into an F′ factor. Since $F^+$ strains can pass the F′ particle into $F^-$ strains, *lac* operon diploids (also called merozygotes, or partial diploids) can be formed. By careful manipulation, various combinations of mutations can be looked at in the diploid state.

### Constitutive Mutants

**Constitutive mutants** are mutants in which the three *lac* operon genes are transcribed at all times—that is, they are not turned off even in the absence of lactose. Inspection of figure 14.3 shows that constitutive production of the enzymes can come about in several ways. A defective repressor, produced by a mutant regulator gene, will not turn the system off, nor will a mutant operator that will no longer bind the normal repressor. The regulator constitutive mutants are designated $i^-$; the operator constitutive mutants are designated $o^c$. Both types of mutants produce the same phenotype: constitutive expression of the three *lac* operon genes.

When a new mutant is isolated, it is possible to determine whether it is caused by a regulator or operator mutation. For example, we can determine the exact location of a mutation on the bacterial chromosome by standard mapping techniques (see chapter 7) or, more recently, by DNA sequencing (see chapter 13). Alternatively, the Jacob and Monod model predicts different modes of action for the two types of mutations. In merozygotes, a constitutive operator mutation affects only the operon it is physically a part of. Operator muta-

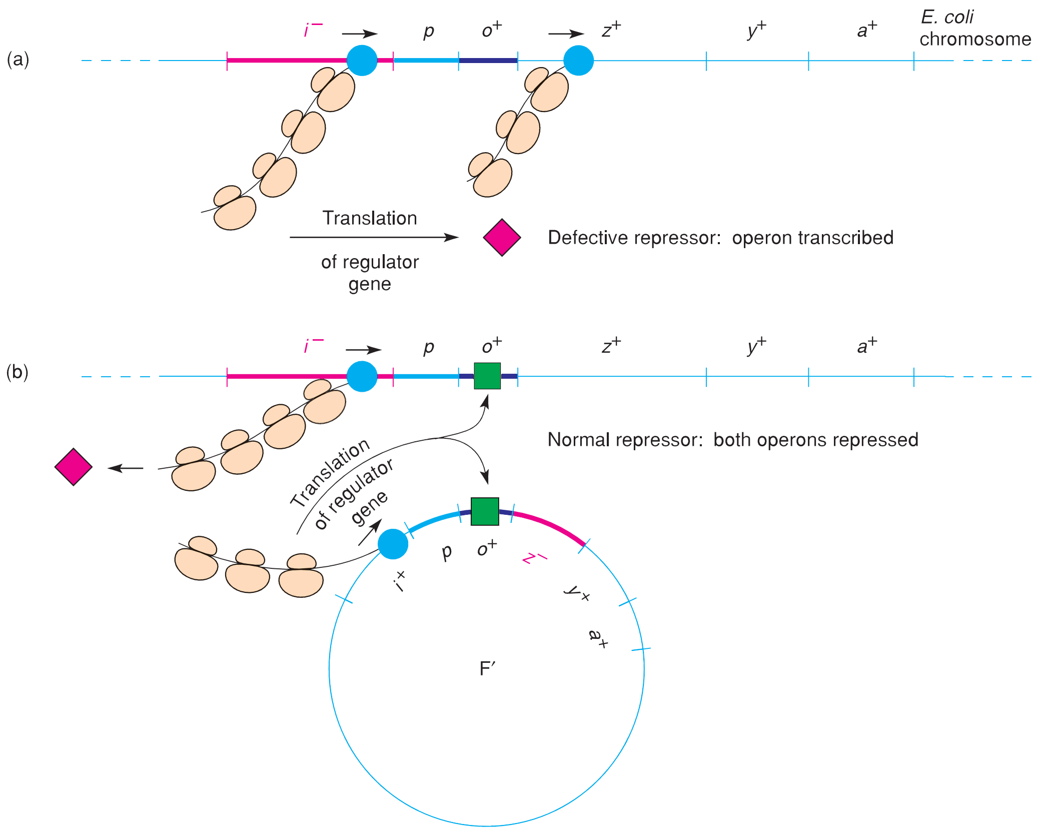

(a)
$i^-$
p
$o^+$
$z^+$
$y^+$
$a^+$
E. coli chromosome
Translation of regulator gene
Defective repressor: operon transcribed
(b)
$i^-$
p
$o^+$
$z^+$
$y^+$
$a^+$
Normal repressor: both operons repressed
Translation of regulator gene
$i^+$
p
$o^+$
$z^-$
$y^+$
$a^+$
F′

**Figure 14.6** (*a*) A *lac* operon in *E. coli* with a mutation of the regulator gene ($i^-$). Transcription and translation of this gene yield a defective repressor; the cell thus has constitutive production of the *lac* operon. In (*b*), the wild-type regulator gene is introduced in an F′ factor; there is both a bacterial chromosome and an F′ factor, each containing a regulator gene. (The F′ operon carries a mutant *z* allele, allowing us to keep track of the transcriptional control of the chromosomal operon only.) In this case, the phenotype is now normal (inducible) because enough repressor is produced by the F′ allele ($i^+$), by transcription and translation, to bind to both operators. RNA polymerase is shown as *solid spheres* on the DNA; the wild-type repressor is shown as a *green square;* the mutant repressor, which cannot bind to the operator, is shown as a *red diamond.*

tions are therefore called ***cis*-dominant.** However, a constitutive *i*-gene mutation, since it works through an altered protein, is recessive to a wild-type regulator gene in the same cell, regardless of which operon (chromosomal or F′ factor) the mutation is on. Constitutive regulator mutations are, therefore, ***trans*-acting.** (If two mutations are on the same piece of DNA, they are in the *cis* configuration. If they are on different pieces of DNA, they are in the *trans* configuration.) *Trans-acting* mutations usually work through a protein product that diffuses through the cytoplasm. *Cis-acting* mutants are changes in recognition sequences on the DNA.

In figure 14.6*a,* the bacterium has a regulator constitutive mutation ($i^-$); the cell has constitutive production of the operon. If the wild-type regulator is introduced in an F′ plasmid (fig. 14.6*b*), the normal (inducible) phenotype is restored because the F′ $i^+$ allele is dominant to the chromosomal mutation—the $i^+$ regulates both the chromosomal and F′ operons. Hence, both operons are inducible. We don’t need to be concerned about the other components of the F′ plasmid because it carries a $z^-$ allele; only the activity of the chromosomal operon will be observed. In figure 14.7*a*, however, the chromosomal operon carries an operator constitutive mutation; the cell also has constitutive production of the operon. When a wild-type operator is introduced into the cell in an F′ plasmid (fig. 14.7*b*), the cell still has the constitutive phenotype because the operator allele on the F′

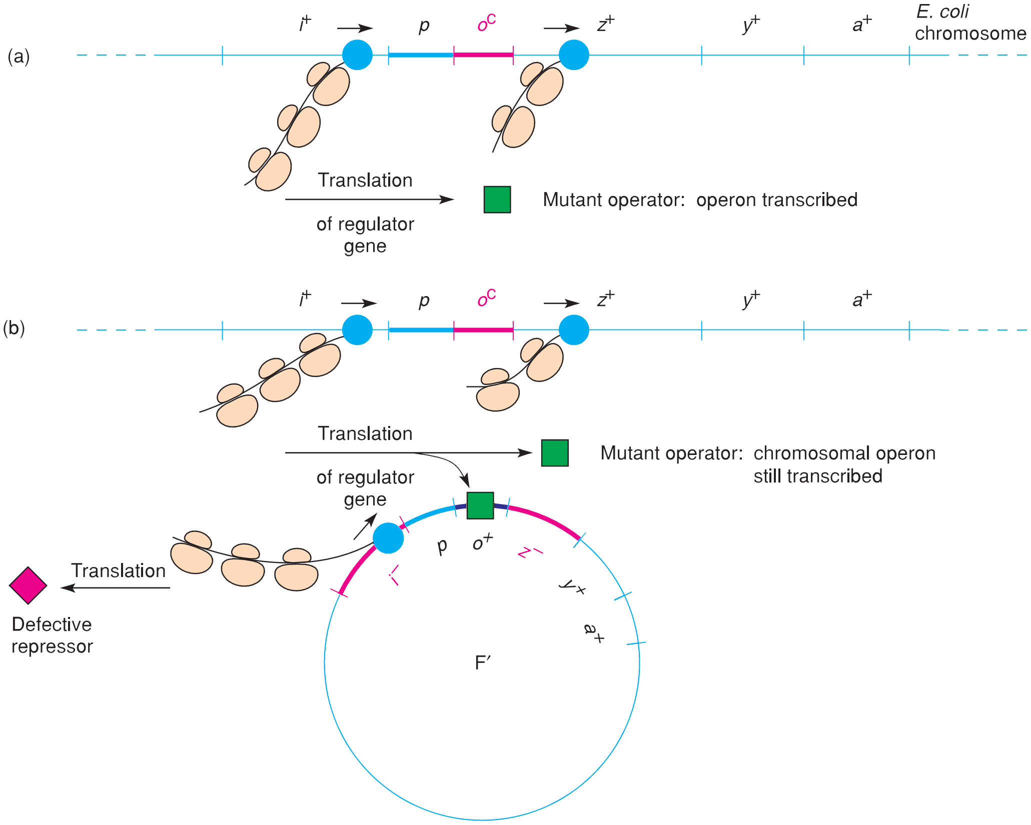

(a)
$i^+$
p
$o^c$
$z^+$
$y^+$
$a^+$
E. coli chromosome
Translation of regulator gene
Mutant operator: operon transcribed
(b)
$i^+$
p
$o^c$
$z^+$
$y^+$
$a^+$
Translation of regulator gene
Mutant operator: chromosomal operon still transcribed
Translation
Defective repressor
$i^-$
p
$o^+$
$z^-$
$y^+$
$a^+$
F′

**Figure 14.7** (*a*) A *lac* operon in *E. coli* with a mutation of the operator ($o^c$). The cell has a constitutive phenotype; the operator cannot bind the wild-type repressor protein, and thus transcription is continuous, even in the absence of lactose. The phenotype is unchanged even when a wild-type operator is introduced into the cell in an F′ factor (*b*); there is both a bacterial chromosome and an F′ factor, each containing an operator. (The F′ operon carries mutant regulator and *z* alleles, allowing us to keep track of the transcriptional control of the chromosomal operon only.) The F′ operator does not change the phenotype of the cell because the wild-type operator exerts no control over the chromosomal operator, which exerts a *cis-dominant* effect; another operator on another operon has no effect. RNA polymerase is shown as *solid spheres* on the DNA; the wild-type repressor is shown as a *green square;* the mutant repressor, which cannot bind to the operator, is shown as a *red diamond.*

plasmid does not control the bacterial operon; the *lac* operon on the bacterial chromosome will be continually transcribed. The chromosomal operon has a *cis-dominant* operator mutation that has a constitutive phenotype. Note, too, that only the bacterial chromosome determines the phenotype because the introduced F′ plasmid has a $z^-$ allele.

### *Other* Lac *Operon Control Mutations*

Other mutations have also been discovered that support the Jacob and Monod operon model. A superrepressed mutation, $i^s$, was located. This mutation represses the operon even in the presence of large quantities of the inducer. Thus, the repressor seems to have lost the ability to recognize the inducer. Basically, the *i*-gene product is acting as a constant repressor rather than as an allosteric protein. In an $i^s/i^+$ merozygote, both operons are repressed because the $i^s$ repressor binds to both operators. Another mutation, $i^Q$, produces much more of the repressor than normal and presumably represents a mutation of the promoter region of the *i* gene.

In 1966, W. Gilbert and B. Müller-Hill isolated the *lac* repressor and thereby provided the final proof of the validity of the model. At about the same time, M. Ptashne and his colleagues isolated the repressor for phage λ operons. Control of gene expression in phage λ is discussed later in this chapter.

Mark Ptashne (1940– ).
(Courtesy of Dr. Mark Ptashne.)

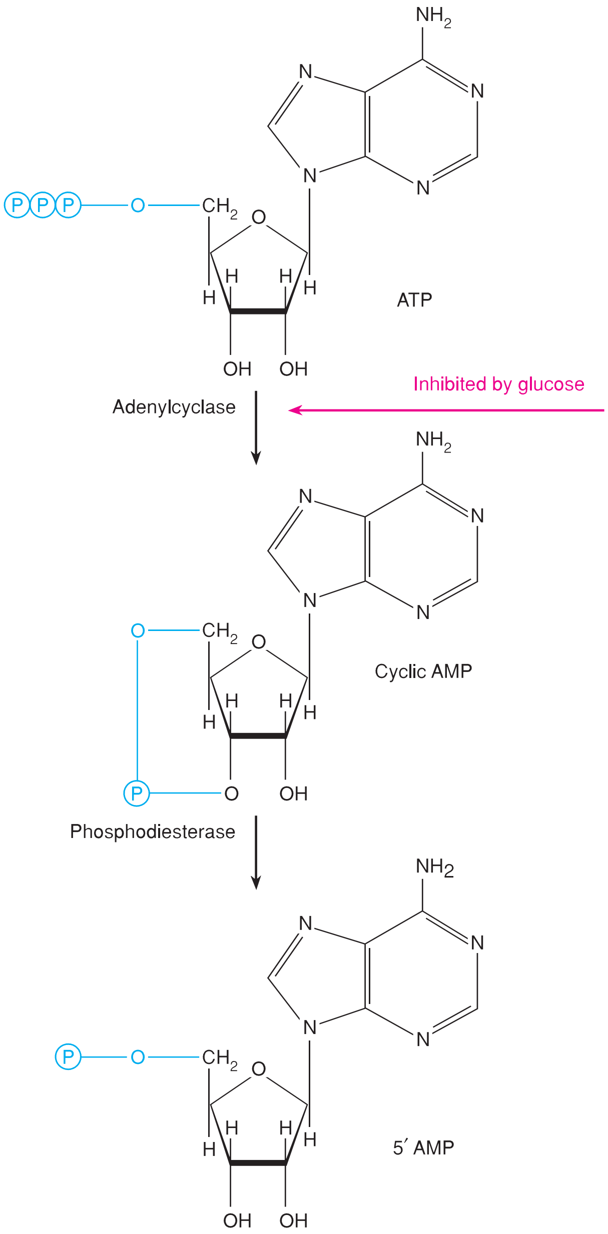


**Figure 14.8** Structure of cyclic AMP (cAMP). Glucose uptake lowers the quantity of cyclic AMP in the cell by inhibiting the enzyme adenylcyclase, which converts ATP to cAMP.

## CATABOLITE REPRESSION

An interesting property of the *lac* operon and other operons that code for enzymes that catabolize certain sugars (e.g., arabinose, galactose) is that they are all repressed in the presence of glucose. That is, glucose is catabolized in preference to other sugars; the mechanism **(catabolite repression)** involves **cyclic AMP** (cAMP; fig. 14.8). In eukaryotes, cAMP acts as a *second messenger,* an intracellular messenger regulated by certain extracellular hormones. Geneticists were surprised to discover cAMP in *E. coli,* where it works in conjunction with another regulatory protein, the **catabolite activator protein (CAP),** to control the transcription of certain operons.

In the absence of glucose, cAMP combines with CAP, and the CAP-cAMP complex binds to a distal part of the promoter of operons with CAP sites (e.g., the *lac* operon; see fig. 14.4). This binding apparently enhances the affinity of RNA polymerase for the promoter, because without the binding of the CAP-cAMP complex to the promoter, the transcription rate is very low. The uptake of glucose by *E. coli* cells causes the loss of cAMP from the cell, probably by inhibiting adenylcyclase (fig. 14.8), and thus lowers the CAP-cAMP level. The transcription rate of operons with CAP sites will, therefore, be reduced (fig. 14.9). The same reduction of transcription rates is noticed in mutant strains of *E. coli* when this part of the distal end of the promoter is deleted. The binding of CAP-cAMP to the CAP site causes the DNA to bend more than 90 degrees (fig. 14.10). This bending, by itself, may enhance transcription, making the DNA more available to RNA polymerase.

In addition, at some point in the process of initiation of transcription, the CAP is in direct contact with RNA polymerase. This was shown by photo cross-linking studies in which the CAP was treated with a cross-linking agent that bound the α subunit of RNA polymerase when irradiated with UV light. For the two proteins to cross-link, they must be in direct contact during the initiation of transcription.

Catabolite repression is an example of positive regulation: Binding of the CAP-cAMP complex at the CAP site enhances the transcription rate of that transcriptional unit. Thus, the *lac* operon is both positively and negatively regulated; the repressor exerts negative control, and the CAP-cAMP complex exerts positive control of transcription.

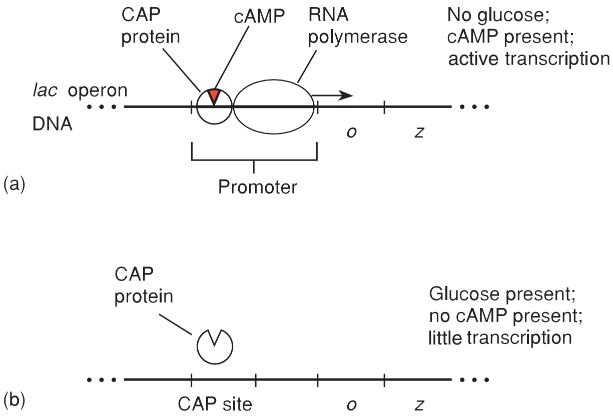

CAP protein
cAMP
RNA polymerase
No glucose; cAMP present; active transcription
*lac* operon
DNA
o
z
(a)
Promoter
CAP protein
Glucose present; no cAMP present; little transcription
(b)
CAP site
o
z

**Figure 14.9** Catabolite repression. When cAMP is present in the cell (no glucose is present), it binds with CAP protein, and together they bind to the CAP site in various sugar-metabolizing operons, such as the *lac* operon shown here. The CAP-cAMP complex enhances the transcription of the operon. When glucose is present, it inhibits the formation of cAMP. Thus no CAP-cAMP complex forms, and transcription of the same operons is reduced.

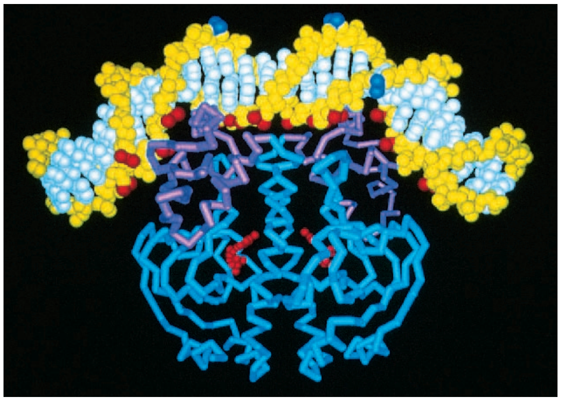


**Figure 14.10** CAP-DNA interaction: model of cap protein and DNA. The cap site has twofold symmetry, like the operator. The cAMP-binding domain is *dark blue,* the DNA-binding domain is *purple,* and the cyclic AMP molecules within the protein are *red.* The DNA sugar-phosphate backbones are shown in *yellow,* the bases in *light blue.* DNA phosphates in *red* (on the double helix) are those whose modification interfere with CAP binding. DNA phosphates in *dark blue* (also on the double helix) are those especially prone to nuclease attack because of the bending of the DNA. (Courtesy of Thomas A. Steitz.)

## *TRP* OPERON (REPRESSIBLE SYSTEM)

The inducible operons are activated when the substrate that is to be catabolized enters the cell. Anabolic operons function in the reverse manner: They are turned off (repressed) when their end product accumulates beyond the needs of the cell. Two entirely different, although not mutually exclusive, mechanisms seem to control the transcription of repressible operons. The first mechanism follows the basic scheme of inducible operons and involves the end product of the pathway. The second mechanism involves secondary structure in messenger RNA transcribed from an attenuator region of the operon.

### *Tryptophan Synthesis*

One of the best-studied repressible systems is the tryptophan, or *trp,* operon in *E. coli.* The *trp* operon contains the five genes that code for the synthesis of the enzymes that build tryptophan, starting with chorismic acid (fig. 14.11). It has a promoter-operator sequence (*p, o*) as well as its own regulator gene (*trpR*).

## Operator Control

In this repressible system, the product of the *trpR* gene, the repressor, is inactive by itself; it does not recognize the operator sequence of the *trp* operon. The repressor only becomes active when it combines with tryptophan. Thus, when tryptophan builds up, enough is available to bind with and activate the repressor. Tryptophan is thus referred to as the **corepressor.** The corepressor-repressor complex then recognizes the operator, binds to it, and prevents transcription by RNA polymerase.

After the available tryptophan in the cell is used up, the diffusion process causes tryptophan to leave the repressor, which then detaches from the *trp* operator. The transcription process no longer is blocked and can proceed normally (the operon is now **derepressed**). Transcription continues until enough of the various enzymes have been synthesized to again produce an excess of tryptophan. Some becomes available to bind to the repressor and make a functional complex, and the operon is again shut off and the process repeated, ensuring that tryptophan is being synthesized as needed (fig. 14.12). This regulation is modified, however, by the existence of the second mechanism for regulating repressible operons—attenuation.

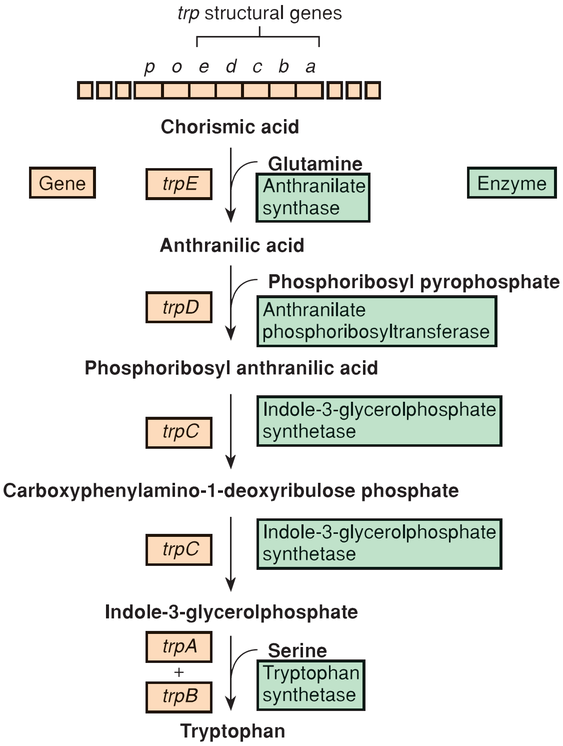

*trp* structural genes
*p* *o* *e* *d* *c* *b* *a*
Chorismic acid
Gene
*trpE*
Glutamine
Anthranilate synthase
Enzyme
Anthranilic acid
*trpD*
Phosphoribosyl pyrophosphate
Anthranilate phosphoribosyltransferase
Phosphoribosyl anthranilic acid
*trpC*
Indole-3-glycerolphosphate synthetase
Carboxyphenylamino-1-deoxyribulose phosphate
*trpC*
Indole-3-glycerolphosphate synthetase
Indole-3-glycerolphosphate
*trpA*
+
*trpB*
Serine
Tryptophan synthetase
Tryptophan

**Figure 14.11** Genes of the tryptophan operon in *E. coli.* The enzymes they produce control the conversion of chorismic acid to tryptophan. The symbol *o* on the chromosome refers to the *trp* operator, which has its own repressor, the product of the *trpR* gene.

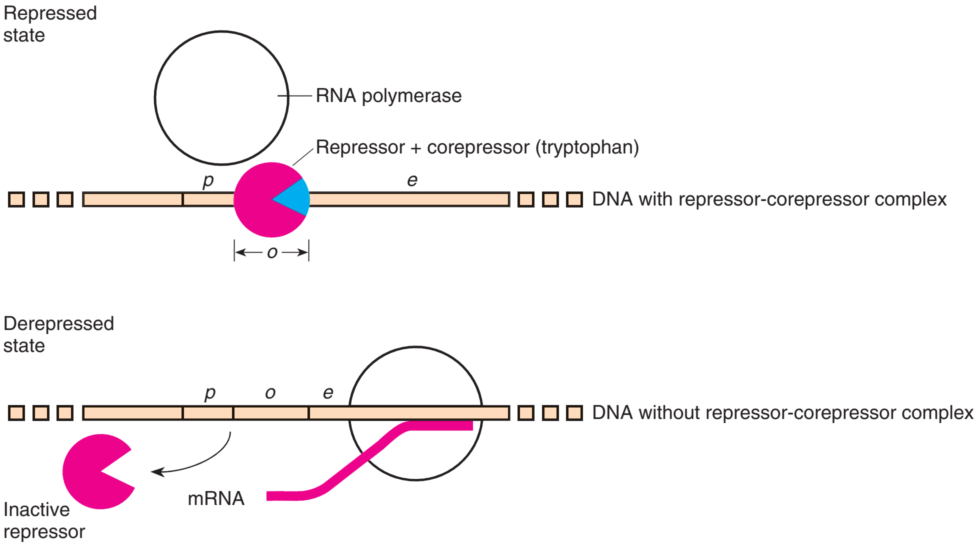

Repressed state
RNA polymerase
Repressor + corepressor (tryptophan)
*p*
*e*
DNA with repressor-corepressor complex
*o*
Derepressed state
*p*
*o*
*e*
DNA without repressor-corepressor complex
mRNA
Inactive repressor

**Figure 14.12** The repressor-corepressor complex binds at the operator and prevents the transcription of the *trp* operon in *E. coli.* Without the corepressor, the repressor cannot bind, and therefore transcription is not prevented. The *blue* wedge is the corepressor (two tryptophan molecules), and the partial *red* circle is the repressor.

## *TRP* OPERON (ATTENUATOR-CONTROLLED SYSTEM)

Details of the second control mechanism of repressible operons have been elucidated primarily by C. Yanofsky and his colleagues, who worked with the tryptophan operon in *E. coli.* This type of operon control, control by an **attenuator region,** has been demonstrated for at least five other amino acid-synthesizing operons, including the leucine and histidine operons. This regulatory mechanism may be the same for most operons involved in the synthesis of an amino acid.

### *Leader Transcript*

Charles Yanofsky (1925– ).
(Courtesy of Dr. Charles Yanofsky.)

In the *trp* operon, an attenuator region lies between the operator and the first structural gene (fig. 14.13). The messenger RNA transcribed from the attenuator region, termed the **leader transcript,** has been sequenced, revealing two surprising and interesting facts. First, four subregions of the messenger RNA have base sequences that are complementary to each other so that three different stem-loop structures can form in the messenger RNA (fig. 14.14). Depending on circumstances, regions 1–2 and 3–4 can form two stem-loop structures, or region 2–3 can form a single stem-loop. When one stem-loop structure is formed, the others are preempted. As

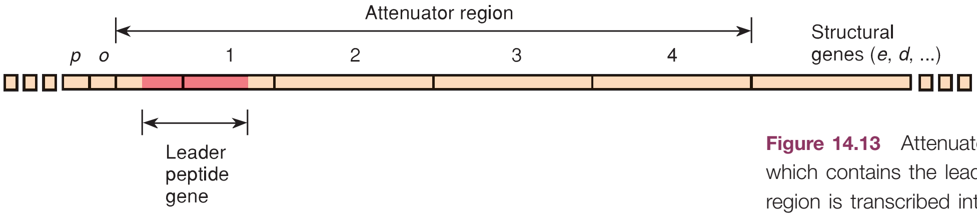


**Figure 14.13** Attenuator region of the *trp* operon, which contains the leader peptide gene (*red*). This region is transcribed into the leader transcript.

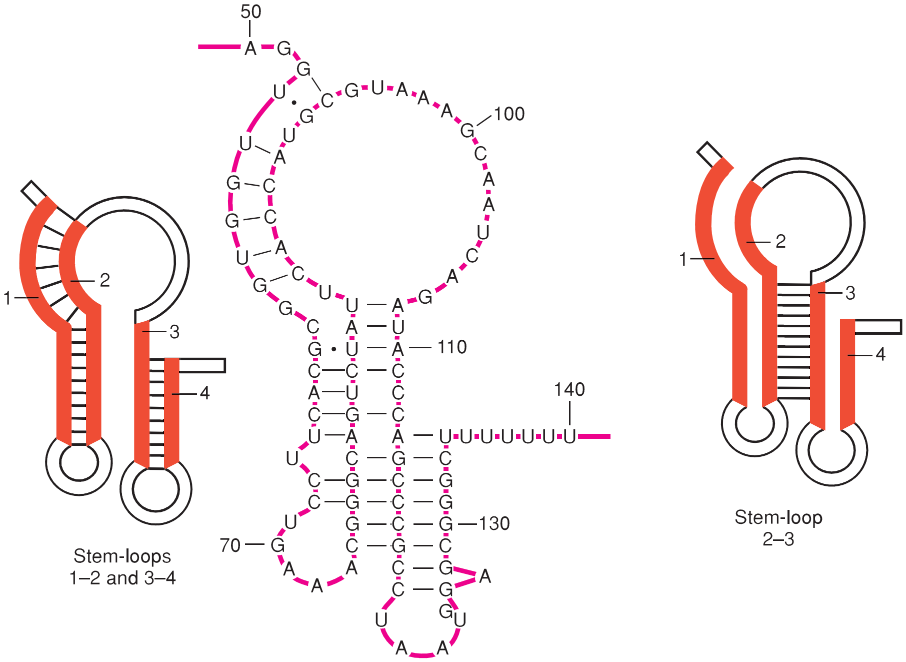


**Figure 14.14** Nucleotide sequence of part of the leader transcript of the *trp* attenuator region (bases 50 to 140). Stem-loops 1–2 and 3–4, or stem-loop 2–3, can form because of complementarity of the nucleotides. All possible base pairings are shown in the middle of the figure. (From D. L. Oxender, et al., "Attenuation in the *Escherichia coli* tryptophan operon: Role of RNA secondary structure involving the tryptophan codon region," *Proceedings of the National Academy of Sciences,* 76:5524–28, 1979. Reprinted by permission.)

we will see, the particular combination of stem-loop structures determines whether transcription continues.

## Leader Peptide Gene

The second fact obtained by sequencing the leader transcript is that there is a small gene coding information for a peptide from bases 27 to 68 (fig. 14.15). The gene for this peptide is referred to as the **leader peptide gene.** It codes for fourteen amino acids, including two adjacent tryptophans. These adjacent tryptophan codons are critically important in attenuator regulation. The proposed mechanism for this regulation follows.

### Excess Tryptophan

Assuming that the operator site is available to RNA polymerase, transcription of the attenuator region will begin. As soon as the 5′ end of the messenger RNA for the leader peptide gene has been transcribed, a ribosome attaches and begins translating this messenger RNA. Depending on the levels of amino acids in the cell, three different outcomes can take place. If the concentration of tryptophan in the cell is such that abundant tryptophanyl-tRNAs exist, translation proceeds down the leader peptide gene. The moving ribosome overlaps regions 1 and 2 of the transcript and allows stem-loop 3–4 to form, as shown in the configuration at the far left of figure 14.16. This stem-loop structure, referred to as the **terminator,** or **attenuator, stem,** causes transcription to be terminated. Note that stem-loop 3–4, the terminator stem, followed by a series of uracil-containing bases, is a rho-independent transcription terminator (see chapter 10). Hence, when existing quantities of tryptophan, in the form of tryptophanyl-tRNA, are adequate for translation of the leader peptide gene, transcription is terminated.

### Tryptophan Starvation

If the quantity of tryptophanyl-tRNA is lowered, the ribosome must wait at the first tryptophan codon until it acquires a Trp-tRNA$^{Trp}$. This is shown in the configuration in the middle part of figure 14.16. The stalled ribosome will

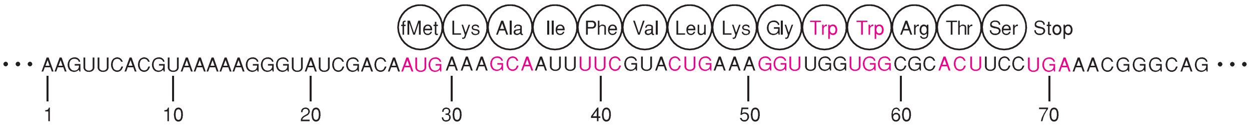

fMet
Lys
Ala
Ile
Phe
Val
Leu
Lys
Gly
Trp
Trp
Arg
Thr
Ser
Stop
• • • AAGUUCACGUAAAAAGGGUAUCGACAAUGAAAGCAAUUUUCGUACUGAAAGGUUGGUGGCGCACUUCCUGAAACGGGCAG • • •
1
10
20
30
40
50
60
70

**Figure 14.15** Base sequence of the *trp* leader transcript and the amino acids these nucleotides code. Note the presence of adjacent tryptophan codons. (From D. L. Oxender, et al., "Attenuation in the *Escherichia coli* tryptophan operon: Role of RNA secondary structure involving the tryptophan codon region," *Proceedings of the National Academy of Sciences,* 76:5524–28, 1979. Reprinted by permission.)

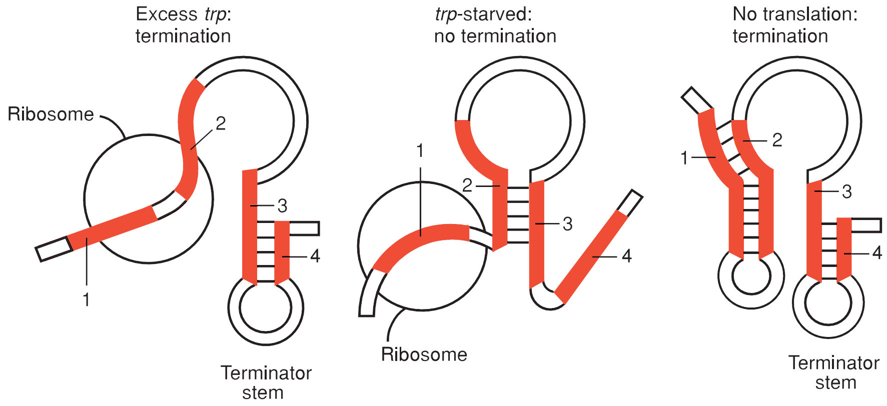

Excess *trp*: termination
Ribosome
2
3
4
1
Terminator stem
*trp*-starved: no termination
1
2
3
4
Ribosome
No translation: termination
1
2
3
4
Terminator stem

**Figure 14.16** Model for attenuation in the *E. coli trp* operon. The *circle* represents the ribosome attempting to translate the leader transcript of figure 14.14. Under conditions of excess tryptophan, the 3–4 stem-loop forms (the terminator stem), terminating transcription. Under conditions of tryptophan starvation, the ribosome is stalled, and stem-loop 2–3 forms, allowing continued transcription. Under general starvation, there is no translation, resulting in the formation of stem-loops 1–2 and 3–4, which again results in the termination of transcription. (From D. L. Oxender, et al., "Attenuation in the *Escherichia coli* tryptophan operon: Role of RNA secondary structure involving the tryptophan codon region," *Proceedings of the National Academy of Sciences,* 76:5524–28, 1979. Reprinted by permission.)

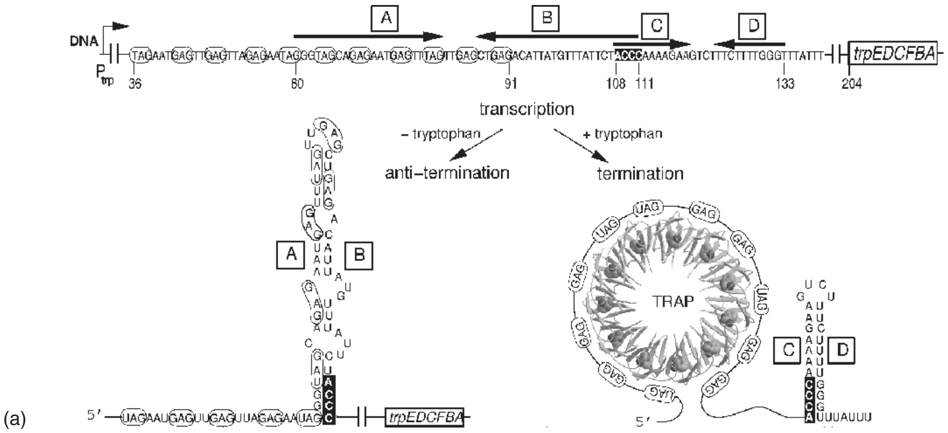
permit stem-loop 2–3 to form, which precludes the formation of the terminator stem-loop (3–4). In this configuration, transcription is not terminated, so that eventually, the whole operon is transcribed and translated, raising the level of tryptophan in the cell. The stem-loop 2–3 structure is referred to as the **preemptor stem.** Note that the preemptor stem is not a rho-independent transcription terminator and thus, without the rho protein present, will not terminate transcription (see chapter 10).

### General Starvation

A final configuration is possible, as shown on the far right in figure 14.16. Here, no ribosome interferes with stem formation and, presumably, stem-loops 1–2 and 3–4 (terminator) form. This configuration also terminates transcription because of the terminator stem. It is believed that this configuration occurs if the ribosome is stalled on the 5′ side of the *trp* codons, which happens when the cell is starved for other amino acids. Presumably, it makes no sense to manufacture tryptophan when other amino acids are in short supply. Hence, the cell can carefully bring up the levels of the various amino acids in the most efficient manner.

## TRAP Control

The tryptophan operon in bacilli such as *Bacillus subtilis* is also controlled by attenuation, but secondary structure in the mRNA transcript is induced by binding not the ribosome, but a ***trp*** **RNA-binding attenuation protein (TRAP).** This protein attaches to the nascent messenger RNA only after the protein binds tryptophan molecules; the result is a terminator stem that forms in the messenger RNA. In the absence of excess tryptophan, TRAP does not bind to the messenger RNA, a preemptor (also called an antiterminator) stem, not the terminator stem, forms, and transcription continues (fig. 14.17*a*). Recently, the



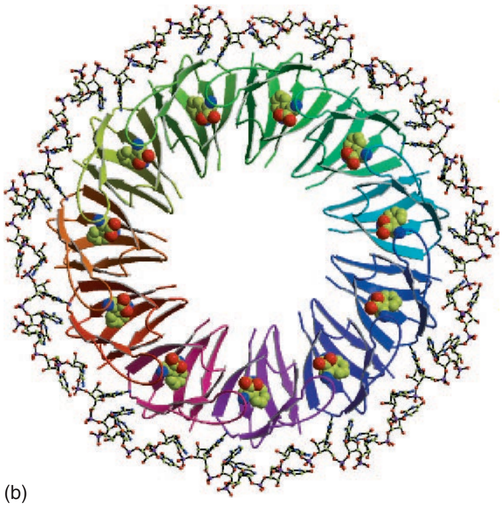


**Figure 14.17** The *trp* operon control by attenuation in *Bacillus subtilis.* (*a*) The top of the figure shows the leader region of the DNA with the two parts of the antitermination stem (A, B) and the termination stem (C, D). The triplets (GAG and TAG in DNA, or GAG and UAG in the messenger RNA) that the *trp* RNA-binding attenuation protein (TRAP) binds to are circled. The label *trpEDCFBA* refers to the structural genes of the *trp* operon. Nucleotides 108–133 (C, D) form the terminator stem, and nucleotides 60–111 (A, B) form the antiterminator stem. The arrows below the boxed letters A–D indicate the inverted repeat sequences forming the stems. Without TRAP, the antiterminator stem forms; with TRAP, the terminator stem forms as TRAP is bound by nucleotides 36–91 of the messenger RNA. Part (*b*) is a close-up of the mRNA (ball-and-stick model) wrapped around TRAP (ribbon diagram with subunits in different colors) bound by tryptophan molecules (spheres). (From Alfred A. Antson, et al., "Structure of the *trp* RNA-binding attenuation protein, TRAP, bound to RNA" in *Nature*, Vol.40, September 16, 1999, fig. 1 p. 234 and fig. 2a p. 237. Reprinted by permission of Macmillan Ltd.)

structure of the protein was worked out; it has eleven symmetrical loops, each of which can bind a tryptophan molecule. When TRAP is bound to tryptophan molecules, it can attach to triplets in the messenger RNA transcript, triplets of GAG or UAG. The TRAP wraps the mRNA around itself, forming an elegant pinwheel (fig. 14.17*b*).

### Redundant Controls

Some amino acid operons are controlled only by attenuation, such as the *his* operon in *E. coli,* in which the leader peptide gene contains seven histidine codons in a row, or the *trp* operon in *B. subtilis.* Redundant control (repression and attenuation) of tryptophan biosynthesis in *E. coli* allows the cell to test both the tryptophan levels (tryptophan is the corepressor) and the tryptophanyl-tRNA levels (in the attenuator control system). The attenuator system also allows the cell to regulate tryptophan synthesis on the basis of the shortage of other amino acids. For example, when there is a shortage of both tryptophan and arginine, operator control allows transcription to begin, but attenuator control terminates transcription because stem-loops 1–2 and 3–4 form (fig. 14.16).

## LYTIC AND LYSOGENIC CYCLES IN PHAGE λ

When a bacteriophage infects a cell, it must express its genes in an orderly fashion; some gene products are needed early in infection, and other products are not needed until late in infection. Early genes usually control phage DNA replication; late genes usually determine phage coat proteins and the lysis of the bacterial cell. A phage is most efficient if it expresses the early genes first and the late genes last in the infection process. Also, temperate phages have the option of entering into lysogeny with the cell; here, too, control processes determine which path is taken. One generalization that holds true for most phages is that their genes are clustered into early and late operons, with separate transcriptional control mechanisms for each.

Phage λ is perhaps the best-studied bacteriophage. It has a chromosome of about 48,500 base pairs. Since it is a temperate phage, it can exist either vegetatively or as a prophage, integrated into the host chromosome. This phage warrants our attention because of the interesting and complex way that its life-cycle choice is determined. It is a model system of operon controls. The complexity results from having two conflicting life-cycle choices.

Briefly, the expression of one of the two life-cycle alternatives, lysogenic or lytic cycles, depends on whether two repressors, CI and Cro, have access to operator sites. The CI repressor acts to favor lysogeny: it represses the lytic cycle. The Cro repressor favors the lytic cycle and represses lysogeny. The operator sites, when bound by either CI or Cro, can either enhance or repress transcription. Other control mechanisms are also involved in determining aspects of the λ life cycle, including antitermination and multiple promoters for the same genes.

### Phage λ Operons

Phage λ (see fig. 7.21) exhibits a complex system of controls of both early and late operons, as well as controls for the decision of lytic infection versus lysogenic integration. The λ genes are grouped into four operons: left, right, late, and repressor (fig. 14.18). The left and right operons contain the genes for DNA replication and recombination and phage integration. The late operon contains the genes that determine phage head and tail proteins and lysis of the host cell. The sequence of events following phage infection is relatively well known.

The map of λ (fig. 14.18) is a circle, but the λ chromosome has two linear stages in its life cycle (fig. 14.19). It is packed within the phage head in one linear form, and it integrates into the host chromosome to form a prophage in another linear form (fig. 14.19). Those two linear forms do not have the same ends (figs. 14.18 and 14.19*b*). The mature DNA, which is packed within the phage heads before lysis of the cells, is flanked by *cos* sites (chapter 13). It results from a break in the circular map between the *A* and *R* loci. The prophage is integrated at the *att* site, and the circular map is thus broken there at integration.

The homologous integration sites on both λ and the *E. coli* chromosome consist of a 15 bp core sequence (called “O” in both), flanked by different sequences on both sides in both the bacterium and the phage (fig. 14.20). In the phage, the region is referred to as POP′, where P and P′ (P for phage) are two different regions flanking the O core on the phage DNA. In the bacterium, the region is called BOB′, where B and B′ (B for bacterium) are two different regions flanking the O core on the *E. coli* chromosome. Integration, which is a part of the lysogenic life cycle, requires the product of the λ *int* gene, a protein known as *integrase,* and is referred to as **site-specific recombination.** Later excision of the prophage, during induction, when the phage leaves the host chromosome to enter the lytic cycle, requires both the integrase and the protein product of the neighboring *xis* gene, *excisionase.*

After infection of the *E. coli* cell by a λ phage, the phage DNA circularizes, using the complementarity of the *cos* sites. Transcription begins, and within a very short time the phage is guided toward either entering the lytic cycle and producing virus progeny or entering

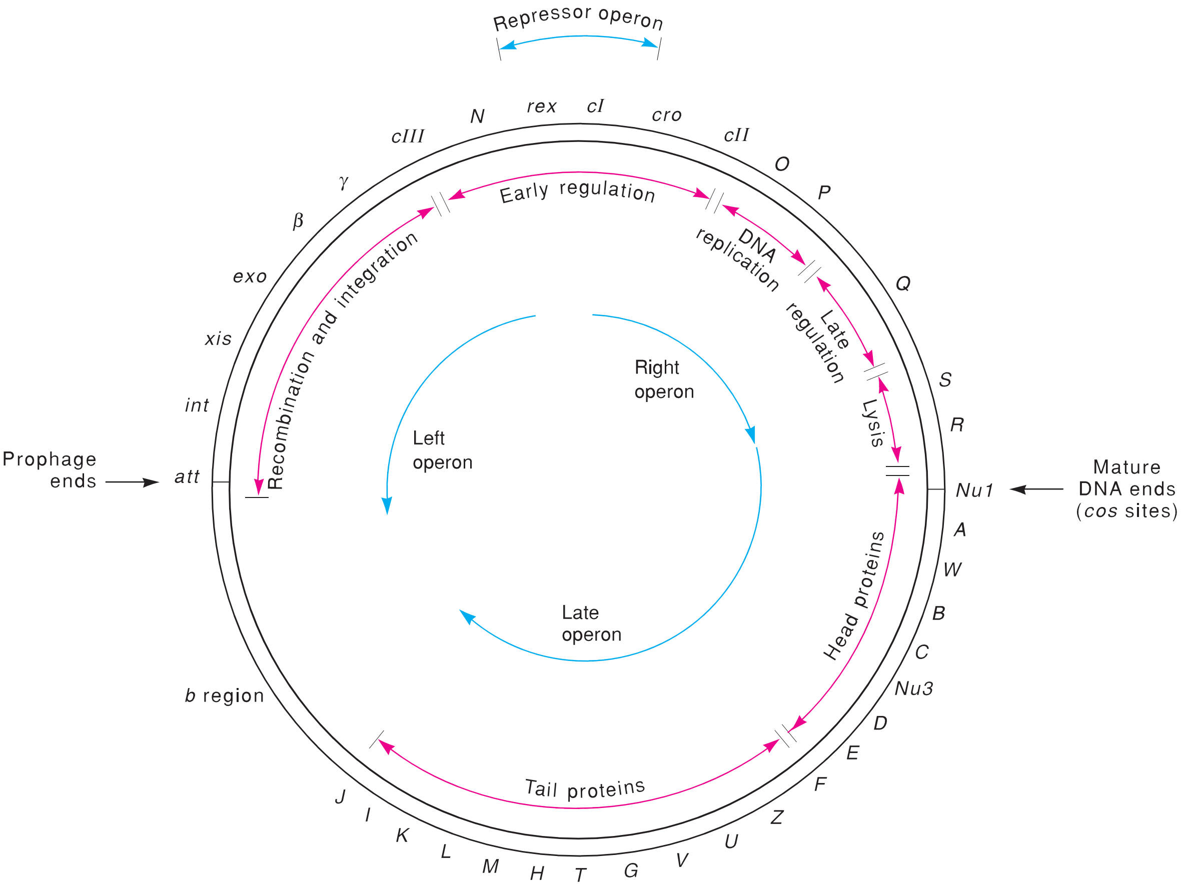
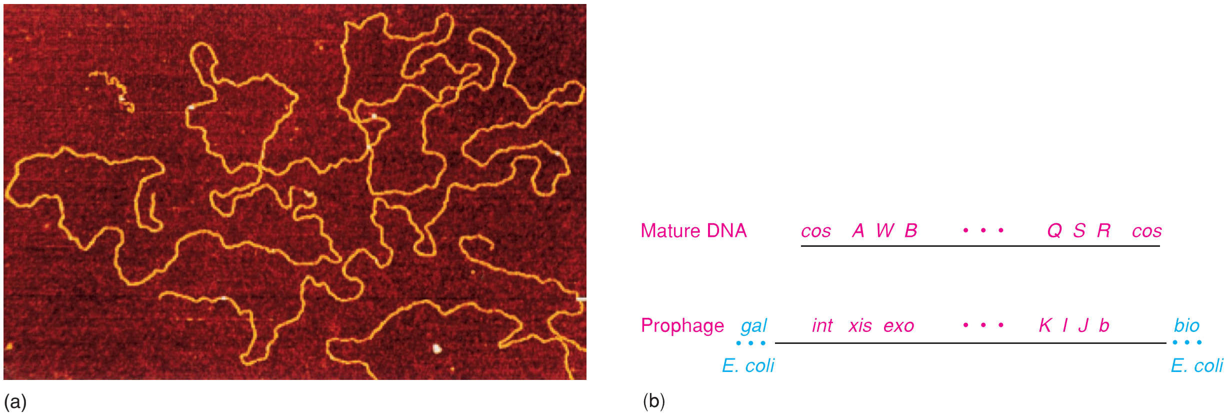

Repressor operon
N
rex
cI
cro
cIII
cII
O
P
γ
β
exo
xis
int
att
Q
S
R
Nu1
A
W
B
C
Nu3
D
E
F
Z
U
V
G
T
H
M
L
K
I
J
b region
Prophage ends
Mature DNA ends (cos sites)
Early regulation
DNA replication
Late regulation
Lysis
Head proteins
Tail proteins
Recombination and integration
Right operon
Left operon
Late operon

**Figure 14.18** Genetic map of phage λ. There are four operons present: the repressor, left, right, and late. The prophage, a linear form integrated into the bacterial chromosome, begins and ends at *att.* The mature phage, another linear form found packed into the phage heads, begins and ends at *Nu1* (*cos* sites).


Mature DNA
cos A W B • • • Q S R cos
Prophage
gal
int xis exo • • • K I J b
bio
E. coli
E. coli

(a)

(b)

**Figure 14.19** The two linear forms of λ phage. (*a*) False color electron micrograph of the λ chromosome, approximately 16 μm in length. This is the linear form of the phage chromosome found within the phage heads. (*b*) The mature linear DNA (found within phage protein coats) is flanked by *cos* sites. The prophage is flanked by *E. coli* DNA (*bio* and *gal* loci). ([*a*] Courtesy of Martin Guthold and Carlos Bustamante, Institute of Molecular Biology and HHMI, University of Oregon.)

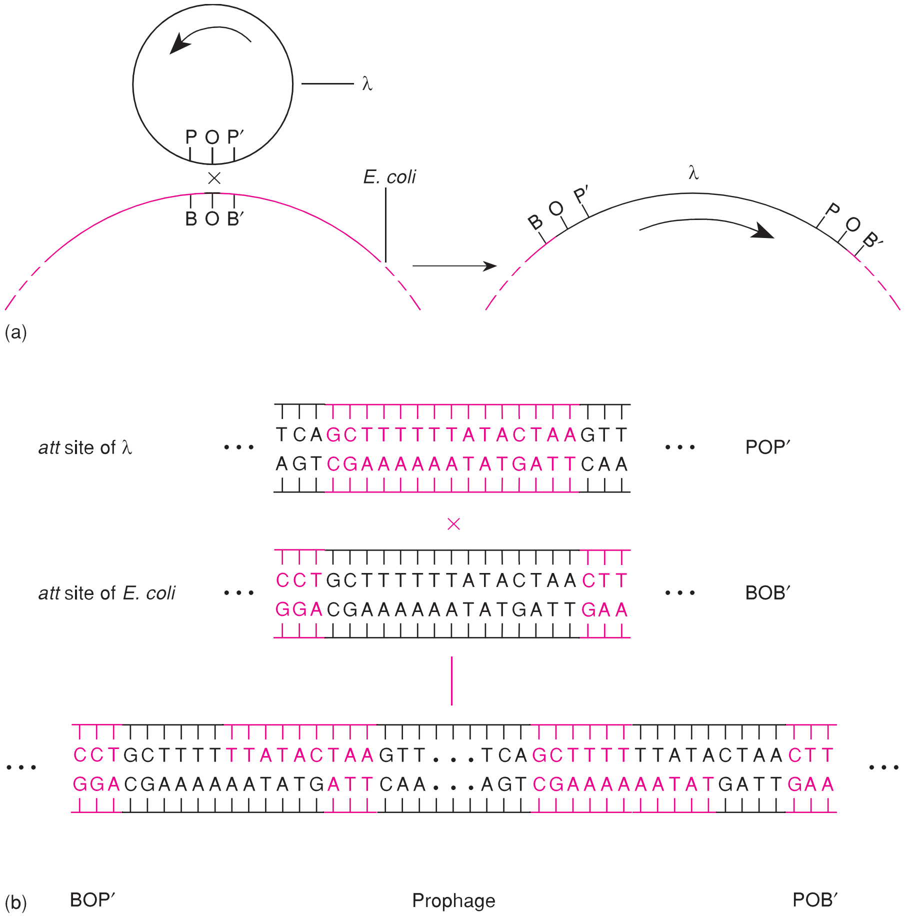

λ
P O P′
×
B O B′
*E. coli*
λ
B O P′
P O B′
(a)
*att* site of λ
TCAGCTTTTTTATACTAAGTT
AGTCGAAAAAATATGATTCAA
POP′
×
*att* site of *E. coli*
CCTGCTTTTTTATACTAACTT
GGACGAAAAAATATGATTGAA
BOB′
CCTGCTTTTTTATACTAAGTT . . . TCAGCTTTTTTATACTAACTT
GGACGAAAAAATATGATTCAA . . . AGTCGAAAAAATATGATTGAA
(b)
BOP′
Prophage
POB′

**Figure 14.20** Integration of the λ phage into the *E. coli* chromosome requires a crossover between the two attach sites, called POP′ (phage) and BOB′ (bacteria). (*a*) General pattern of this site-specific attachment. (*b*) Nucleotide sequences of the various components.

the lysogenic cycle and integrating into the host chromosome. What events lead up to this "decision" on which path to take?

## Early and Late Transcription

When the phage first infects an *E. coli* cell, transcription of the left and right operons begins at the left ($p_L$) and right ($p_R$) promoters, respectively: The *N* (left) and *cro* (right) genes are transcribed (fig. 14.21) and then translated into their respective proteins. Transcription then stops on both operons at rho-dependent terminators ($t_{R1}$, $t_{L1}$). Transcription cannot continue until the protein product of the *N* gene is produced. This protein is called an **antiterminator protein.** When it binds at sites upstream from the terminators, called *nutL* and *nutR* (*nut* stands for *N* *ut*ilization; *L* and *R* stand for left and right), the polymerase reads through the terminators and continues on to transcribe the left and right operons. (Although it is not completely clear why antitermination

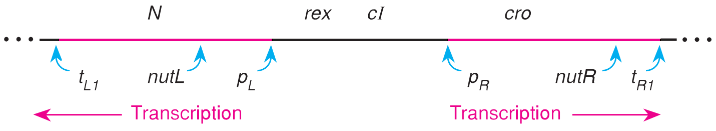

N
rex
cI
cro
$t_{L1}$
nutL
$p_L$
$p_R$
nutR
$t_{R1}$
Transcription
Transcription

**Figure 14.21** Transcription begins at the left and right promoters ($p_L$, $p_R$) and proceeds to the left and right terminators ($t_{L1}$, $t_{R1}$). Transcription continues through these terminators when the protein product of the *N* gene binds to the *nutL* and *nutR* sites.

has evolved here, it seems to give the phage better control over the timing of events.)

Transcription then continues along the left and right operons through the *cII* and *cIII* genes (see fig. 14.18). Later, if the lytic response is followed, the *Q* gene, which codes for a second antiterminator protein in the right operon, has the same effect on the late operon as the *N* gene did on the two early operons: Without the *Q*-gene product, transcription of the late operon proceeds about two hundred nucleotides and then terminates. With the *Q*-gene product, the late operon is transcribed. Hence, in phage λ, proteins that allow RNA polymerase to proceed past termination signals mediate general control of transcription. If only the previously described events were to transpire, the lytic cycle pathway would always be followed. However, a complex series of events can also take place in the repressor region that may lead to a "decision" to follow the lysogenic cycle instead.

## Repressor Transcription

The *cIII*-protein product inhibits a host cell protease, called FtsH, that would break down the *cII*-gene product. The *cII*-gene product binds at two promoters, enhancing their availability to RNA polymerase, just as the CAP-cAMP product enhances the transcription of the *lac* operon. The *cII* protein binds at the promoters for *cI* transcription and for *int* transcription (fig. 14.22). At this point, the phage can still "choose" between either the lytic or the lysogenic cycles. Integrase (the product of the *int* gene) and *cI* (repressor) proteins are now produced, favoring lysogeny, as well as the *cro*-gene product, the *antirepressor,* which is a repressor of *cI* and therefore favors the lytic pathway. (*Cro* stands for *c*ontrol of *r*epression and *o*ther things; the *c* of *cI,* the repressor, stands for "clear," which is the appearance of λ plaques that have *cI* mutations. These mutants can only undergo lysis without the possibility of lysogeny. Normal λ infections produce turbid plaques, accounted for by lysogenic bacterial growth within the plaques.) We now focus further on the repressor region with its operators and promoters.

## Maintenance of Repression

The *cI* gene, with the aid of the *cII*-gene product, is transcribed from a promoter known as $p_{RE}$, the *RE* standing for *r*epression *e*stablishment (fig. 14.23). Once *cI* is transcribed, it is translated into a protein called the λ *repressor,* which interacts at the left and right operators, $o_L$ and $o_R$, of the left and right operons. When these operators are bound by *cI* protein, transcription of the left and right operons (and therefore also the late operon) ceases. There are several ramifications of the repression. First, lysogeny can be initiated because the *int* gene has been transcribed at the early stage of infection. Second, since CII and CIII are no longer being synthesized, *cI* transcription from the $p_{RE}$ promoter stops. However, *cI* can still be transcribed because there is a second promoter, $p_{RM}$ (*RM* stands for *r*epression *m*aintenance), that allows low levels of transcription of the *cI* gene.

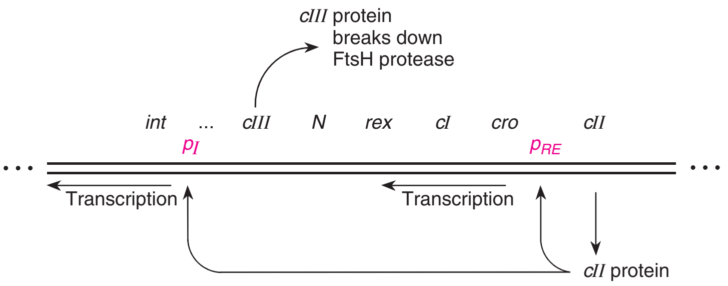


**Figure 14.22** The *cII*-gene product of phage λ binds to the *cI* promoter ($p_{RE}$) and the *int* promoter ($p_I$), enhancing transcription of those genes. The *cIII* protein breaks down the FtsH protease that would normally break down the *cII* protein.

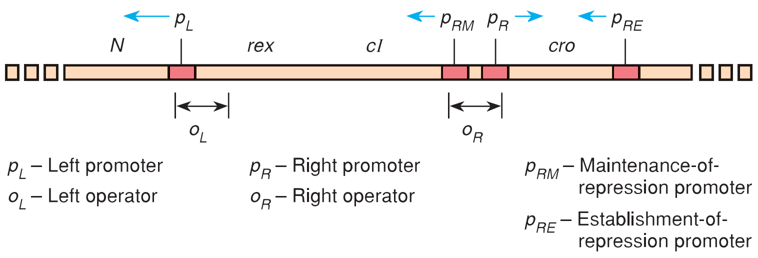


**Figure 14.23** Early regulation region of phage λ. Two promoters, $p_{RE}$ and $p_{RM}$, transcribe the *cI* and *rex* genes. The left operator overlaps the left promoter, and the right operator overlaps both the right promoter and the maintenance-of-repression promoter.

The *cI* gene can further control its own concentration in the cell. When the right and left operators were sequenced, each was discovered to have three sites of repressor recognition (fig. 14.24). On the right operator, for example, the right-most site ($o_{R1}$) was found to be most efficient at binding repressor. When repressor was bound at this site, the right operon was repressed, and transcription of *cI* was enhanced (again, in a way similar to enhancement of transcription by the binding of CAP-cAMP at the CAP site in the *lac* operon). Excess repressor, when present, however, was also bound by the other two sites within $o_R$. The foregoing process results in the repression of the *cI* gene itself. Hence, maintenance levels of CI can be kept within very narrow limits.

A third ramification of repression is the prevention of superinfection. That is, bacteria lysogenic for λ phage are protected from further infection by other λ phages because repressor is already present in the cell. Thus, the excess of repressor controls new invading λ phages. (We say that bacterial cells lysogenic for phage λ are immune from infection by additional λ phage.) These bacteria are also protected from infection by T4 phage with *rII* mutants. The *rex*-gene product, the product of the other gene in the repressor operon, controls this protection.

The promoters for maintenance and establishment of repression differ markedly in their control of repressor gene expression. When $p_{RE}$ is active, a very high level of repressor is present, whereas $p_{RM}$ produces only a low level of repressor. The level of repressor is due to the length of the leader RNA transcribed on the 5′ side of the *cI* gene. The $p_{RE}$ promoter transcribes a very long leader RNA and is very efficient at translation of the *cI* region. In contrast, the $p_{RM}$ promoter begins transcription at the initiation codon of the protein. This leaderless messenger RNA is translated very inefficiently into CI.

The λ repressor is a dimer of two identical subunits (fig. 14.25). Each subunit is composed of two domains, or

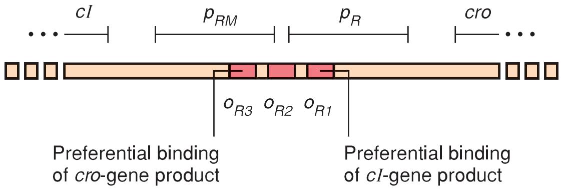

cI
$p_{RM}$
$p_R$
cro
$o_{R3}$ $o_{R2}$ $o_{R1}$
Preferential binding of *cro*-gene product
Preferential binding of *cI*-gene product

**Figure 14.24** The right operator on the phage λ chromosome overlaps the $p_{RM}$ and $p_R$ promoters. There are three repressor recognition sites within the operator: $o_{R1}$, $o_{R2}$, and $o_{R3}$. Preferential binding by the Cro repressor to $o_{R3}$ and the CI repressor to $o_{R1}$ determines whether transcription occurs to the left or the right.

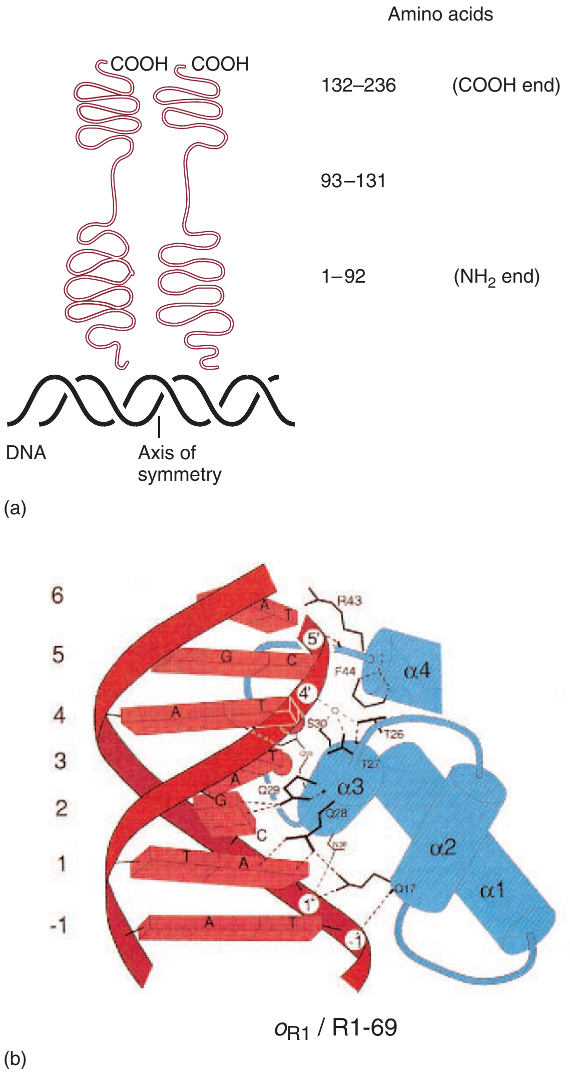


**Figure 14.25** The λ repressor. (*a*) The λ repressor is a dimer with each subunit having helical amino- and carboxyl-terminal ends. The helical structure of each amino-terminal end binds in the major DNA groove. (*b*) Diagram of the interaction of amino acid residues 1–69 (*blue*) with $o_{R1}$ (*red*) in the closely related phage 434. *Dashed lines* are hydrogen bonds. Numbers −1 to 6 and 4′ and 5′ are phosphate numbers. Amino acids are designated by the single-letter code (fig. 11.1). Small *red circles* are water molecules. (From D. W. Rodgers and S. C. Harrison, "The complex between phage 434 repressor DNA-binding domain and operator site $O_{R3}$: Structural differences between consensus and non-consensus half-sites," *Structure,* 1:227–40, Dec. 15, 1993. © Current Biology Ltd.)

"ends." The carboxyl- and amino-terminal ends are separated by a relatively open region, susceptible to protease attack. The alpha-helical regions of the amino-terminal ends interdigitate into the major groove of the DNA to locate the specific sequences making up the left and right operator sequences. As described earlier for the *lac* operator, $o_{R1}$, $o_{R2}$, and $o_{R3}$ each have twofold symmetry.

The binding of the λ repressor in $o_{R1}$ enhances the binding of another molecule of repressor into $o_{R2}$. Together, they enhance $p_{RM}$ transcription, presumably through contact with RNA polymerase. The repressors also block $p_R$ transcription (see fig. 14.24).

## Lysogenic Versus Lytic Response

We have described the mechanism by which λ establishes lysogeny. How then does λ turn toward the lytic cycle? Here, control is exerted by the *cro*-gene product, another repressor molecule that works at the left and right operators in a manner antagonistic to the way the CI repressor works. In other words, using the right operator as an example, *cro*-gene product binds preferentially to the leftmost of the three sites within $o_R$ and represses *cI* but enhances the transcription of *cro* (see fig. 14.24).

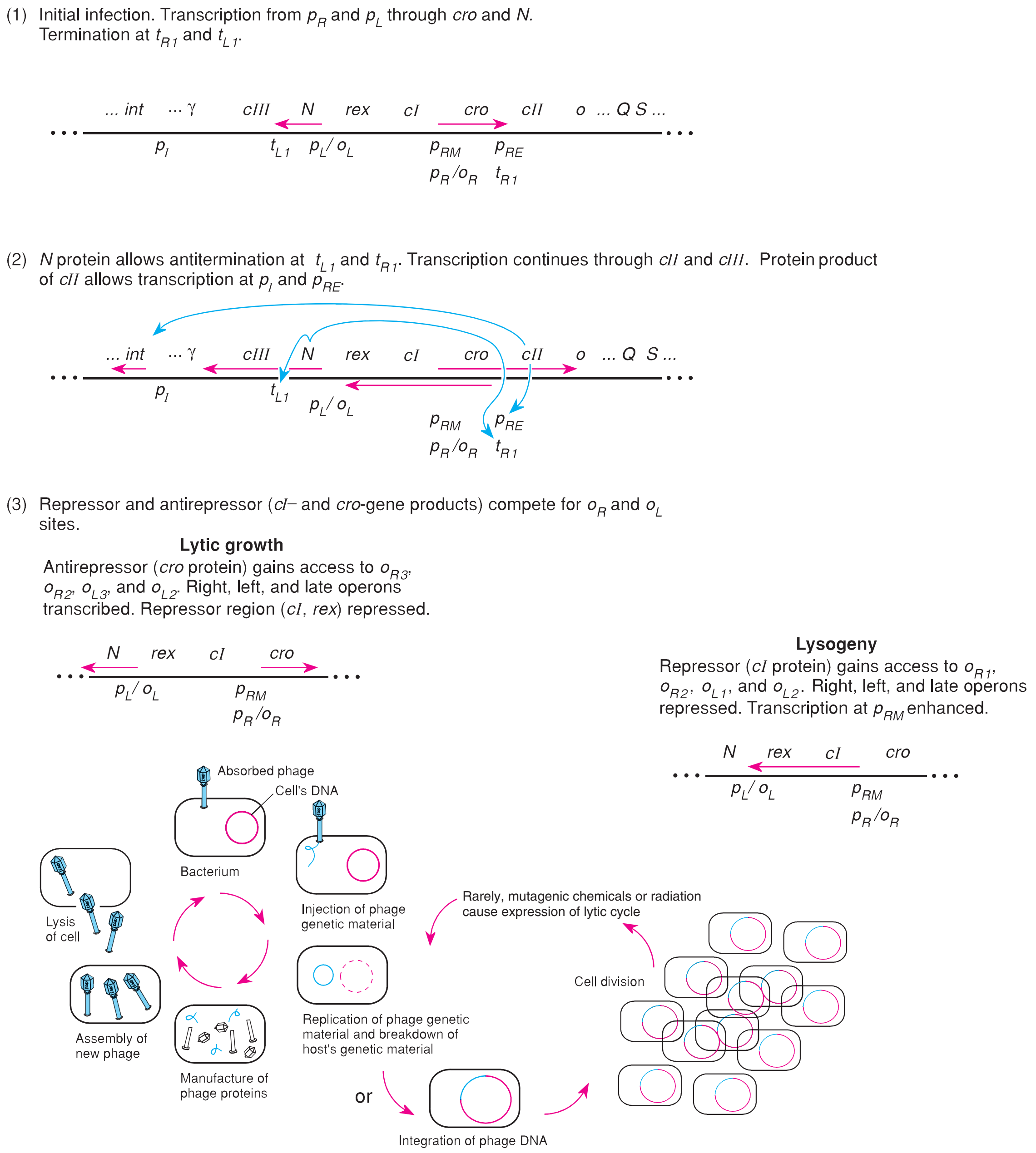
The *cro*-gene product can direct the cell toward a lytic response if it occupies the $o_R$ and $o_L$ sites before the λ repressor, or if the λ repressor is removed. From the point of view of phage λ, when would be a good time for the CI repressor to be removed? Thinking in evolutionary terms, we would expect that a prophage might be at an advantage if it left a host's chromosome and began the lytic cycle when it "sensed" damage to the host. In fact, one of the best ways to induce a prophage to enter the lytic cycle is to direct ultraviolet (UV) light at the host bacterium. (Actually, this was how lysogeny was discovered, by French geneticist André Lwoff.) UV light causes damage to DNA and induces several repair systems. One, called SOS repair (see chapter 12), makes use of the protein product of the *recA* gene. Among the activities of this enzyme is to cleave the λ repressor in the susceptible region between domains. The cleaved repressor falls free of the DNA, making the operator sites available for the *cro*-gene product. The lytic cycle then follows.

Initially, however, when the phage first infects an *E. coli* cell, the "decision" for lytic versus lysogenic growth is probably determined by the *cII*-gene product. This protein, as we mentioned, is susceptible to a bacterial protease, which, in turn, is an indicator of cell growth. When *E. coli* growth is limited, its proteases tend to be limited, a circumstance that would favor lysogeny for the phage. It is the *cII* protein that, when active, favors lysogeny and when inactive favors the lytic cycle. Thus, under active bacterial growth, the *cII* protein is more readily destroyed, it thus fails to enhance *cI* transcription, and lysis follows. When bacteria are not growing actively, the *cII* protein is not readily destroyed, it enhances *cI* transcription, and lysogeny results. Thus, under initial infection, the choice between lysogeny or the lytic cycle depends primarily on the *cII* protein, which gauges the health and activity of the host. After lysogeny is established, it can be reversed by processes that inactivate the *cI* protein, indicating genetic damage to the bacterium (the SOS response) or an abundance of other hosts in the environment (zygotic induction, see chapter 7). In zygotic induction, the lytic cycle is induced during conjugation, presumably when an Hfr cell sends a copy of the λ prophage into an $F^-$ cell. At that point, without repressor present, the prophage can reassess whether to continue lysogeny or enter the lytic cycle.

Not all the details regarding the CI-Cro competition are known, but an understanding of the relationship of lytic and lysogenic life cycles and the nature of DNA-protein recognition has emerged (fig. 14.26 and table 14.1).

**Table 14.1** **Elements in Phage λ Infection**

| | |
|---|---|
| **Gene Products** | |
| *cI* | Repressor protein whose function favors lysogeny |
| *cII* | Enhances transcription at the $p_I$ and $p_{RE}$ promoters |
| *cIII* | Inhibits the *FtsH* protease |
| *cro* | Antirepressor protein that favors lytic cycle |
| *N* | Antiterminator acting at *nutR* and *nutL* |
| *rex* | Protects bacterium from infection by T4 *rII* mutants |
| *int* | Integrase for prophage integration |
| *Q* | Antiterminator of late operon |
| *FtsH* | Bacterial protease that degrades *cII* protein |
| **Promoters of** | |
| $p_R$ | Right operon |
| $p_L$ | Left operon |
| $p_{RE}$ | Establishment of repression at repressor region |
| $p_{RM}$ | Maintenance of repression at repressor region |
| $p_R$ | Late operon |
| $p_I$ | *int* gene |
| **Terminators** | |
| $t_{R1}$ | Terminates after *cro* gene |
| $t_{L1}$ | Terminates after *N* gene |
| **Antiterminators** | |
| *nutR* | In *cro* gene |
| *nutL* | In *N* gene |



(1) Initial infection. Transcription from $p_R$ and $p_L$ through *cro* and *N*. Termination at $t_{R1}$ and $t_{L1}$.

(2) *N* protein allows antitermination at $t_{L1}$ and $t_{R1}$. Transcription continues through *cII* and *cIII*. Protein product of *cII* allows transcription at $p_I$ and $p_{RE}$.

(3) Repressor and antirepressor (*cI*– and *cro*-gene products) compete for $o_R$ and $o_L$ sites.

**Lytic growth**

Antirepressor (*cro* protein) gains access to $o_{R3}$, $o_{R2}$, $o_{L3}$, and $o_{L2}$. Right, left, and late operons transcribed. Repressor region (*cI*, *rex*) repressed.

**Lysogeny**

Repressor (*cI* protein) gains access to $o_{R1}$, $o_{R2}$, $o_{L1}$, and $o_{L2}$. Right, left, and late operons repressed. Transcription at $p_{RM}$ enhanced.

... int
... γ
cIII
N
rex
cI
cro
cII
o
... Q S ...
$p_I$
$t_{L1}$
$p_L/o_L$
$p_{RM}$
$p_{RE}$
$p_R/o_R$
$t_{R1}$
Absorbed phage
Cell's DNA
Bacterium
Injection of phage genetic material
Lysis of cell
Assembly of new phage
Manufacture of phage proteins
Replication of phage genetic material and breakdown of host's genetic material
or
Integration of phage DNA
Rarely, mutagenic chemicals or radiation cause expression of lytic cycle
Cell division

**Figure 14.26** Summary of regulation of phage λ life cycles. (*1*) In the initial infection, transcription begins in *cro* and *N* but terminates shortly thereafter at left and right terminators. (*2*) The product of the *N* gene allows transcription through the initial terminators; in essence, all genes can now be transcribed. (*3*) Lysogeny will occur if the *cI* protein gains access to the right and left operators; the lytic cycle will prevail if the *cro*-gene product gains access to those two operators.

# TRANSPOSABLE GENETIC ELEMENTS

Up until this point, we have thought of the genome in fairly conservative terms. If we map a gene today, we expect to see it in the same place tomorrow. However, our discovery of mobile genetic elements has modified that view to some extent. We now know that some segments of the genome can move readily from one place to another. Their moving can have an effect on the phenotype of the organism, primarily at the transcriptional level. We thus begin our discussion of mobile genetic elements here, and we conclude it in chapter 16, because mobile elements also affect the phenotypes of eukaryotes.

## *IS Elements*

**Transposable genetic elements, transposons,** or even *jumping genes,* are regions of the genome that can move from one place to another. In some cases, transposition is *conservative:* the transposons move without copying themselves. They are liberated from the donor site by double-strand breaks in the DNA. In other cases, transposition is *replicative:* a copy of the transposon is inserted while the original stays in place. This mechanism involves only single-strand breaks of the DNA at the donor site.

Barbara McClintock first discovered transposable elements in corn in the 1940s (see chapter 16); they were discovered in prokaryotes in 1967, where they first showed up as **polar mutants** in the galactose operon of *E. coli.* No genes of the operon were expressed past the point of the polar mutation. This effect was explained by assuming that the transposon brought with it a transcription stop signal. The presence of an inserted piece of DNA in these polar mutants was verified by heteroduplex analysis (fig. 14.27).

The first transposable elements discovered in bacteria were called **insertion sequences** or **IS elements.** It turns out that these are the simplest transposons. The IS elements consist of a central region of about 700 to 1,500 base pairs surrounded by an inverted repeat of about 10 to 30 base pairs, the numbers depending on the specific IS element. Presumably, the inverted repeats signal the transposing enzyme that it is at the ends of the IS element. The central region of the IS element contains a gene or genes for the transposing event (usually genes for transposase and resolvase enzymes); the relatively small IS elements carry no bacterial genes (fig. 14.28).

The target site that the transposable element moves to is not a specific sequence, as with the *att* site of λ. It becomes a direct repeat flanking the IS element only after insertion, giving rise to a model of insertion (fig. 14.29). The target site is cut in a staggered fashion, leaving single-stranded ends. The IS element is then inserted between the single-stranded ends. Repair processes convert the two single-stranded tails to double-stranded segments and, hence, to direct flanking repeats. When DNA is sequenced, the pattern of a direct flanking repeat surrounding an inverted repeat, with a segment in the middle, signals the existence of a transposable element. Currently, we know of more than fifteen families, including a total of over five hundred known members, of IS elements.

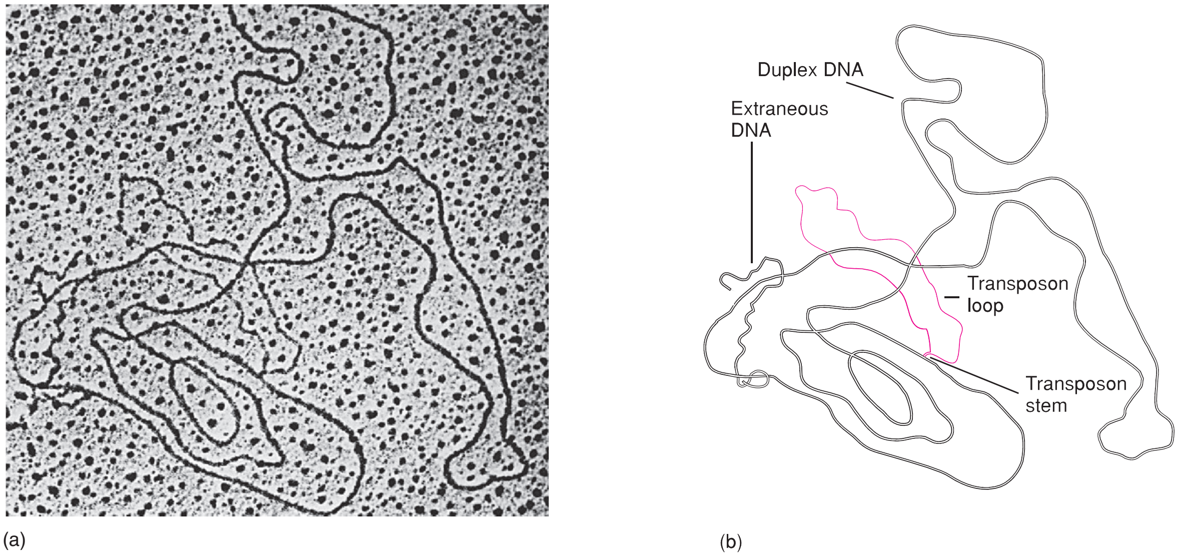


Figure 14.27 Heteroduplex analysis revealing a transposon. (*a*) Two plasmids were hybridized, one with and one without a transposon. (*b*) The transposon is seen as a single-stranded loop (*red*); it has nothing to pair with in the heteroduplex. ([*a*] Courtesy of Richard P. Novick, M.D.)

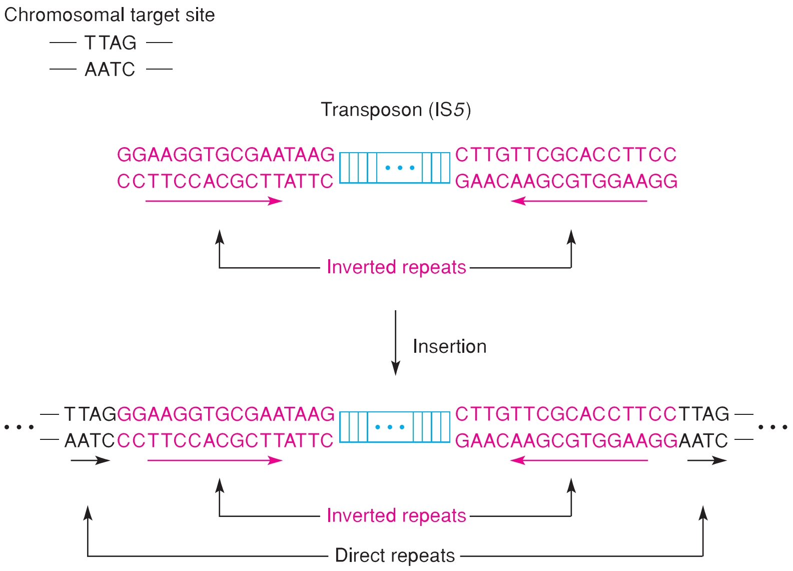

Chromosomal target site
— TTAG —
— AATC —
Transposon (IS*5*)
GGAAGGTGCGAATAAG
CCTTCCACGCTTATTC
CTTGTTCGCACCTTCC
GAACAAGCGTGGAAGG
Inverted repeats
Insertion
. . . — TTAGGGAAGGTGCGAATAAG
— AATCCCTTCCACGCTTATTC
CTTGTTCGCACCTTCCTTAG — . . .
GAACAAGCGTGGAAGGAATC —
Inverted repeats
Direct repeats

**Figure 14.28** An IS element (IS*5*) inserted into a target site in a bacterial chromosome creates a direct repeat on either side of the IS element. The explanation is shown in figure 14.29. An inverted repeat (*red*) is seen as the same sequence read inwards from outside on the upper strand (*left*) and the lower strand (*right*). (Reprinted with permission from *Nature,* Vol. 297, M. Kroger and G. Hoborn, "Structural Analysis of Insertion Sequence IS5." Copyright © 1982 Macmillan Magazines Limited.)

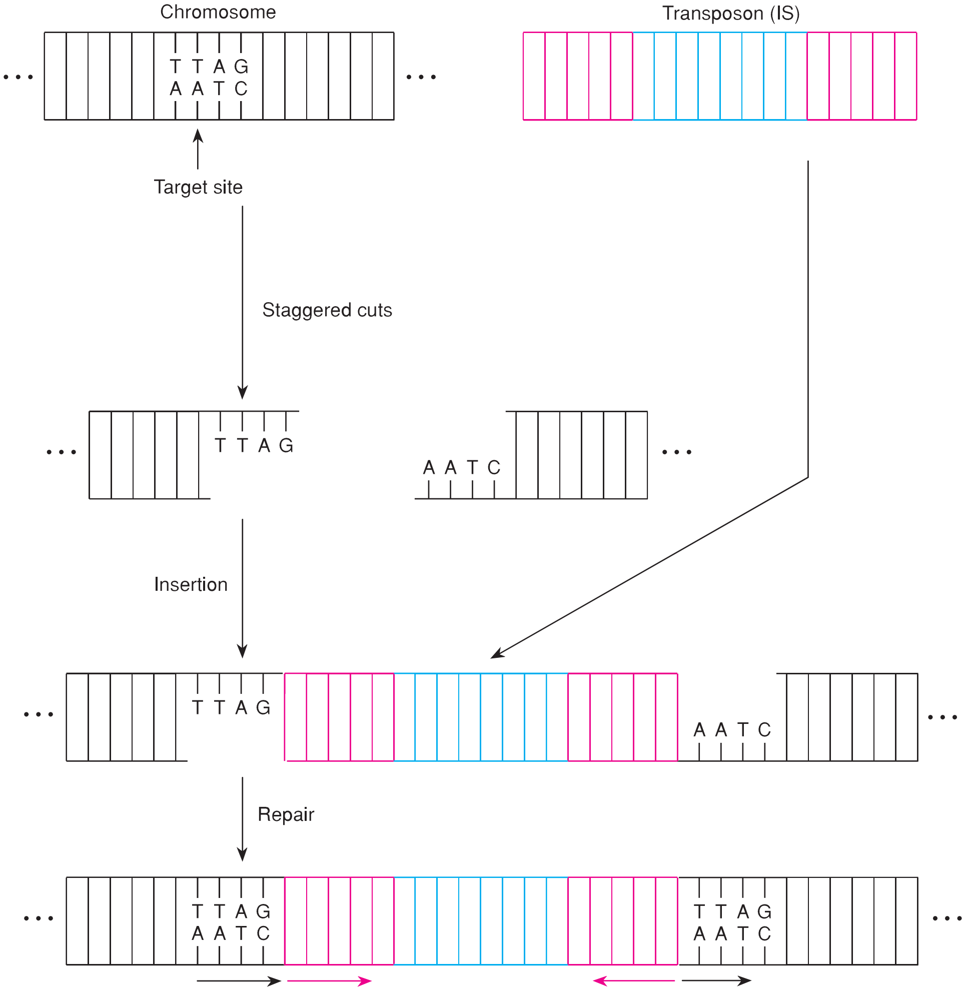

Chromosome
Transposon (IS)
T T A G
A A T C
Target site
Staggered cuts
T T A G
A A T C
Insertion
T T A G
A A T C
Repair
T T A G
A A T C
T T A G
A A T C

**Figure 14.29** Insertion of an IS element (IS*5* of fig. 14.28) results in a direct flanking repeat surrounding the transposon in the host chromosome. This occurs because the insertion takes place at a point in which a staggered cut is made in the host DNA, leaving complementary regions on either side of the transposon. Repair replication results in two copies of the flanking sequence.

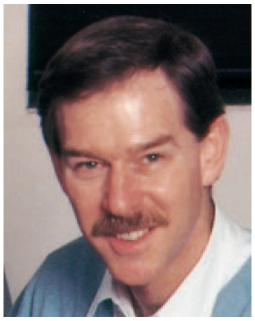
## *Composite Transposons*

After the discovery of IS elements, a more complex type of transposable element, a **composite transposon,** was discovered. A composite transposon consists of a central region surrounded by two IS elements. The central region usually contains bacterial genes, frequently antibiotic resistance loci. For example the composite transposon Tn*10* contains the genes for transposase and resolvase, as well as the bacterial gene for β-lactamase, which confers resistance to ampicillin (fig. 14.30). Arrangements of composite transposons can vary quite a bit. The IS elements at the two ends can be identical or different; they can be in the same or different orientations; they can be similar to known IS elements or different from any freely existing IS elements. In the latter case, they are called IS-like elements.

Two IS elements can transpose virtually any region between them. In fact, composite transposons most likely came into being when two IS elements became located near each other. We can see this very clearly in a simple experiment. In figure 14.31, there is a small plasmid constructed with transposon Tn*10* in it. The "reverse" transposon, consisting of the two IS elements and the plasmid genes, or the normal transposon, could each transpose.

## *Mechanism of Transposition*

Transposition can come about by several mechanisms; however, it does not use the normal recombination machinery of the cell (see chapter 12). One model, by J. Shapiro, explains the fact that many transposons in the



J. A. Shapiro (1943– ). (Courtesy of Dr. J. A. Shapiro.)

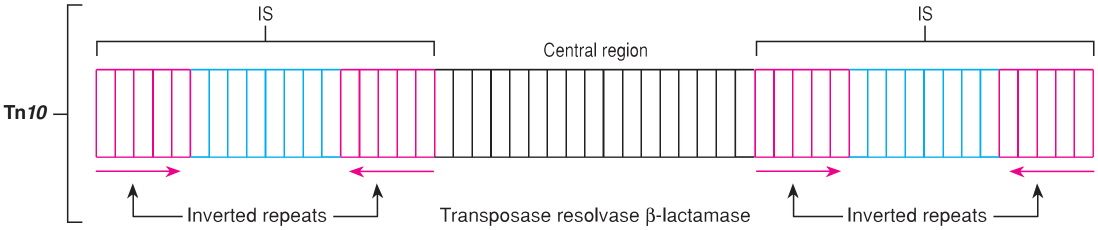


**Figure 14.30** A composite transposon consists of a central region flanked by two IS elements. Transposon Tn*10* contains the transposase and resolvase enzyme genes as well as the bacterial gene β-lactamase, which protects the cell from the antibiotic ampicillin.

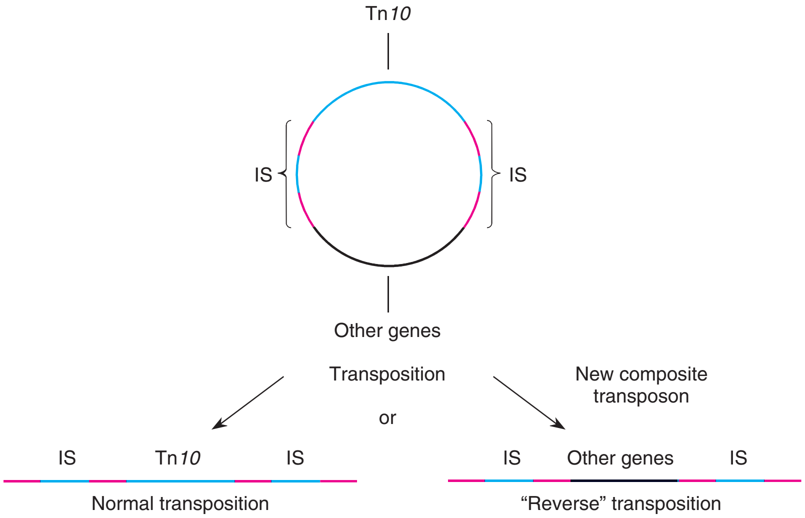


**Figure 14.31** Two IS elements in a plasmid can transpose either of the two regions between them. In the case shown, either the Tn*10* transposon or the "reverse" transposon ("other genes") is transposed.

process of transposition go through a **cointegrate** state (fig. 14.32), in which there is a fusion of two elements. During the process of transposition (in this case from one plasmid to another), an intermediate cointegrate stage is formed, made up of both plasmids and two copies of the transposon. Then, through a process called *resolution,* the cointegrate is reduced back to the two original plasmids, each now containing a copy of the transposon.

A diagram of Shapiro's mechanism appears in figure 14.33. At first, staggered cuts are made in the donor and recipient DNA molecules (fig. 14.33*a* and *b*). Then nonhomologous ends are joined so that only one strand of the transposon connects them (fig. 14.33*c*). This process is presumably controlled by the transposon-coded *transposase* enzyme. Repair-DNA replication now takes place to fill in the single-stranded segments. The result is a cointegrate of the two plasmids with two copies of the transposon. The last step is a recombination event at a homologous site within the two transposons. This is catalyzed by a *resolvase* enzyme, which resolves the cointegrate into the original two plasmids, each with a copy of the transposon (fig. 14.33*e*).

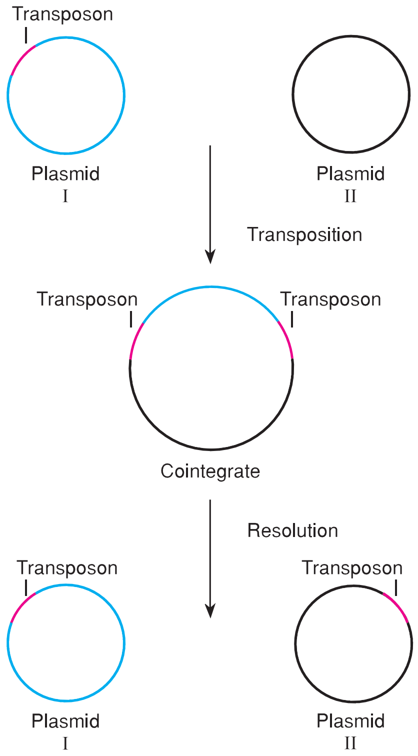

Transposon
Plasmid I
Plasmid II
Transposition
Transposon
Transposon
Cointegrate
Resolution
Transposon
Transposon
Plasmid I
Plasmid II

**Figure 14.32** Transposition frequently goes through an intermediate cointegrate stage. In this case, the transposon is copied from one plasmid to another, with an intermediate stage consisting of a single large plasmid.

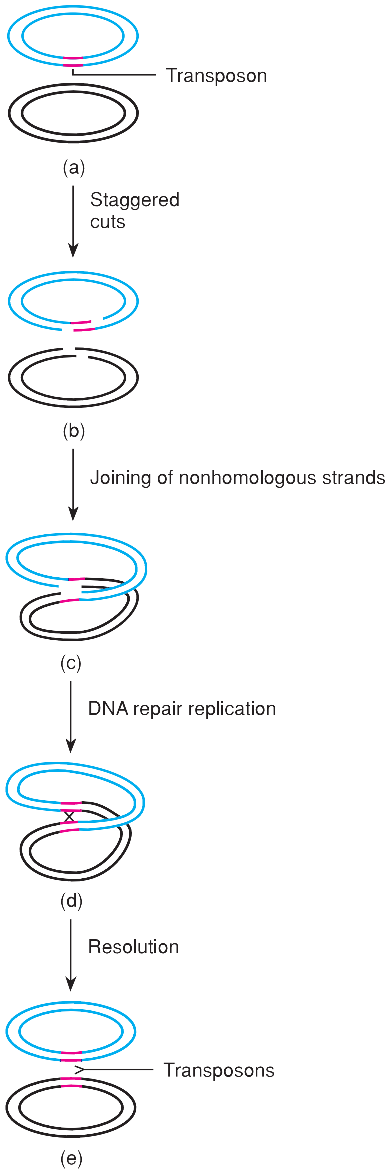
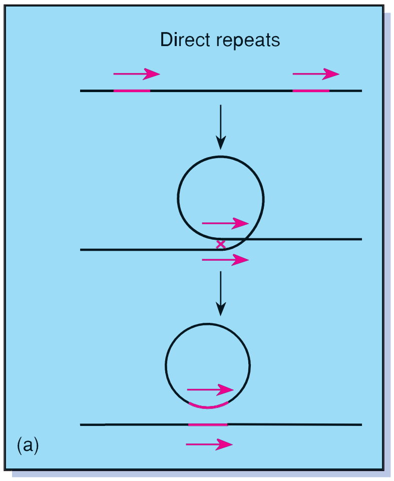

Transposon
(a)
Staggered cuts
(b)
Joining of nonhomologous strands
(c)
DNA repair replication
(d)
Resolution
Transposons
(e)

**Figure 14.33** The Shapiro mechanism of transposition. Staggered cuts are made at the site of transposon insertion and at either side of the transposon itself (*a* and *b*). Nonhomologous single strands join, resulting in two single-stranded copies of the transposon in the cointegrate (*c*). Repair replication of these single strands produces two copies of the transposon (*d*). A crossover at the transposon resolves the cointegrate into two plasmids, each with a copy of the transposon (*e*).

### *Phenotypic and Genotypic Effects of Transposition*

Transposition can have several effects on the phenotype and genotype of an organism. If transposition takes place into a gene or its promoter, it can disrupt the expression of that gene. Depending on the orientation of a transposon, it can prevent the expression of genes. A transposon can also cause deletions and inversions.

Direct repeats on a chromosome can come about, for example, by the sequential transposition of the same IS or transposon, in the same orientation. Pairing followed by recombination results in a deletion of the section between the repeats (fig. 14.34). In the case of inverted repeats, pairing followed by recombination results in an inversion of the section between the repeats.



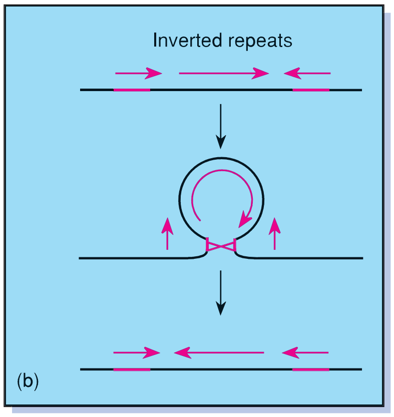


**Figure 14.34** Pairing and recombination in DNA repeats. (*a*) Direct repeats can result in deletion (in the form of a *circle*) due to a single crossover. (*b*) Inverted repeats can result in an inversion of the region between the repeats due to a crossover.

A well-known case of transposon orientation controlling a phenotype in bacteria occurs in *Salmonella typhimurium.* The flagella of this bacterium occur in two types. Any particular bacterium has either type 1 or type 2 flagella (called phase 1 or phase 2 flagella). The difference is in the flagellin protein the flagella are composed of. Phase 1 flagella are determined by the *H1* gene and phase 2 flagella are determined by the *H2* gene. The change from one phase to another occurs at a rate of about $10^{-4}$ per cell division. After extensive genetic analysis, the following scheme was suggested and later verified using recombinant DNA techniques.

The *H1* and *H2* genes are at separate locations on the bacterial chromosome (fig. 14.35). *H2* is part of an operon that also contains the *rH1* gene, the repressor of *H1.* The promoter of this operon lies within a transposon upstream of the operon. When the promoter is in the proper orientation, the *H2* operon is expressed, resulting in phase 2 flagella. The *rH1* gene product represses the *H1* gene (fig. 14.35*a*). If the inverted repeat ends of the transposon undergo recombination, the transposon is inverted (see fig. 14.34), moving the promoter into an incorrect orientation for the transcription of the *H2* operon. No *H1* repressor is made, so the *H1* gene is expressed (fig. 14.35*b*).

As N. Kleckner has summarized, transposons can have marked effects on the phenotype by their actions in transposition and by the fact that they may carry genes valuable to the cell. However, they can also exist without any noticeable consequences. This fact has led some evolutionary geneticists to suggest that transposons are an evolutionary accident that, once created, are self-maintaining. Since they may exist without a noticeable benefit to the host's phenotype, transposons have been

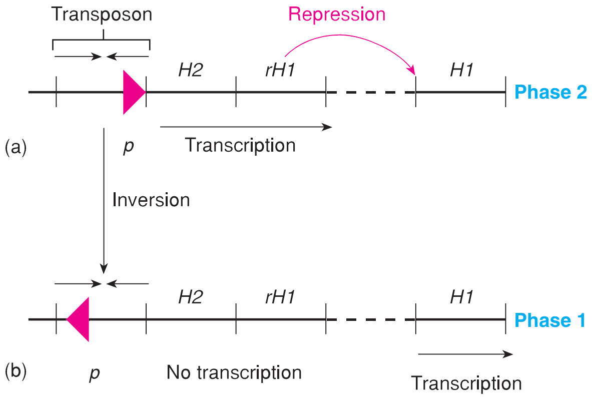


**Figure 14.35** Arrangement of flagellin genes on the *Salmonella* chromosome. The promoter (*p*) is within a transposon. In one orientation (*a*), the *H2* operon is transcribed, which results in *H2* flagellin and *rH1* protein, the repressor of the *H1* gene. In the second orientation (*b*), the *H2* operon is not transcribed, resulting in uninhibited transcription of the *H1* gene.

Nancy Kleckner (1947– ). (Courtesy of Nancy Kleckner. Photo by Stu Rosner.)

referred to as **selfish DNA.** In recent theoretical and experimental studies, however, some scientists have suggested that transposons improve the evolutionary fitness of the bacteria that have them (see chapter 20).

## OTHER TRANSCRIPTIONAL CONTROL SYSTEMS

### Transcription Factors

#### Phage T4

Phage T4, a relatively large phage with seventy-three genes, has transcription controlled by particular RNA polymerase specificity factors. Like phage λ, phage T4 has early, middle, and late genes, genes that need to be expressed in a particular order. Early T4 genes have promoters whose specificity of recognition depends on the sigma factor of the host ($\sigma^{70}$ of *E. coli*). Two products of early phage genes are the AsiA and MotA proteins. AsiA binds to the −35 recognition region of $\sigma^{70}$, preventing transcription from both host genes and early phage genes. AsiA is thus called an **anti-sigma factor,** a protein that interferes with a sigma factor. Middle phage promoters do not have −35 recognition regions but do bind MotA. Host RNA polymerase bound with the $\sigma^{70}$-AsiA complex recognizes these promoters. Finally, late phage genes have promoters that depend on the phage-encoded sigma factor $\sigma^{gp55}$. Some proteins are needed both early and late in the infection process; they are specified by genes that have promoters recognized by different specificity factors.

#### Heat Shock Proteins

A response to elevated temperature, found in both prokaryotes and eukaryotes, is the production of heat shock proteins (see chapter 10). In *E. coli,* elevated temperatures cause the general shutdown of protein synthesis concomitant with the appearance of at least seventeen heat shock proteins. These proteins help protect the cell against the consequences of elevated temperature; some are molecular chaperones (see chapter 11). The production of these proteins is the direct result of the gene product of the *htpR* gene, which codes for $\sigma^{32}$. The normal sigma factor is $\sigma^{70}$, the product of the *rpoD* gene; the heat shock genes have promoters recognized by $\sigma^{32}$ rather than $\sigma^{70}$. Heat causes the *htpR* gene to become active, as well as stabilizing $\sigma^{32}$. From DNA sequence data, the difference in promoters between normal genes and heat shock genes seems to lie in the −10 consensus sequence (Pribnow box). In normal genes, it is TATAAT; in heat shock protein genes, it is CCCCATXT, in which *X* is any base.

### Promoter Efficiency

In addition to the mechanisms previously described, there are other ways to regulate the transcription of messenger RNA. One is to control the efficiency of various processes. For example, we know that the promoter sequence of different genes in *E. coli* differs. Since the affinity for RNA polymerase is different for the different sequences, the rate of initiation of transcription for these genes also varies. The more efficient promoters are transcribed at a greater rate than the less efficient promoters. An example is the promoter of the *i* gene of the *lac* operon. This promoter is for a constitutive gene that usually produces only about one messenger RNA per cell cycle. However, mutants of the promoter sequence are known that produce up to fifty messenger RNAs per cell cycle. Here, then, the transcriptional rate is controlled by the efficiency of the promoter in binding RNA polymerase. Efficiency can be controlled by the direct sequence of nucleotides (i.e., differences from the consensus sequence) or by the distance between consensus regions. For example, promoters vary in the number of bases between the −35 and −10 sequences. Seventeen seems to be the optimal number of bases separating the two. Presumably, more or fewer than seventeen reduces the efficiency of transcription.

## TRANSLATIONAL CONTROL

When considering control of gene expression, it is important to remember that all control mechanisms are aimed at exerting an influence on either the amount, or the activity, of the gene product. Therefore, in addition to transcriptional controls, which influence the amount of messenger RNA produced, there are also translational controls affecting how efficiently the messenger RNA is translated. (Attenuator control—see fig. 14.16—can also be viewed as translational control because the environment is tested by translation even though attenuation results in the cessation of transcription.) In prokaryotes, translational control is of lesser importance than transcriptional control for two reasons. First, messenger RNAs are extremely unstable; with a lifetime of only about two minutes, there is little room for controlling the rates of translation of existing messenger RNAs because they simply do not last very long. Second, although there are some indications of translational control in prokaryotes, such control is inefficient—energy is wasted synthesizing messenger RNAs that may never be used.

Translational control can be exerted on a gene if the gene occurs distally from the promoter in a polycistronic operon. The genes that are transcribed last appear to be translated at a lower rate than the genes transcribed first. The three *lac* operon genes, for instance, are translated roughly in a ratio of 10:5:2. This ratio is due to the polarity of the translation process. That is, in prokaryotes, translation is directly tied to transcription—a messenger RNA can have ribosomes attached to it well before transcription ends. Thus, genes at the beginning of the operon are available for translation before genes at the end. In addition, exonucleases seem to degrade messenger RNA more efficiently from the 3′ end. Presumably, natural selection has ordered the genes within operons so that those producing enzymes needed in greater quantities will be at the beginning of an operon.

Translation can also be regulated by RNA-RNA hybridization. RNA complementary to the 5′ end of a messenger RNA can prevent the translation of that messenger RNA. The regulating RNA is called **antisense RNA.** In figure 14.36, the messenger RNA from the *ompF* gene in *E. coli* is prevented from being translated by complementary base pair binding with an antisense RNA called *micF* RNA (*mic* stands for *m*RNA-*i*nterfering *c*omplementary RNA). The *ompF* gene codes for a membrane component called a *porin,* which, as the name suggests, provides pores in the cell membrane for transport of materials. Surprisingly, a second porin gene, *ompC,* seems to be the source of the *micF* RNA. Transcription of the opposite DNA strand (the one not normally transcribed) near the promoter of the *ompC* gene yields the antisense RNA. One porin gene thus seems to regulate the expression of another porin gene, for reasons that are not completely understood. Antisense RNA has also been implicated in such phenomena as the control of plasmid number and the control of transposon Tn*10* transposition. Control by antisense RNA is a fertile field for gene therapy because antisense RNA can be artificially synthesized and then injected into eukaryotic cells.

A third translational control mechanism consists of the efficiency with which the messenger RNA binds to the ribosome. This is related to some extent to the sequence of nucleotides at the 5′ end of the messenger RNA that is complementary to the 3′ end of the 16S ribosomal RNA segment in the ribosome (the Shine-Dalgarno sequence). Variations from the consensus sequences demonstrate different efficiencies of binding and, therefore, the initiation of translation occurs at different rates.

The redundancy in the genetic code can also play a part in translational control of some proteins since different transfer RNAs occur in the cell in different quantities. Genes with abundant protein products may have codons that specify the more common transfer RNAs, a concept called **codon preference.** In other words, certain codons are preferred; they specify transfer RNAs that are abundant. Genes that code for proteins not needed in abundance could have several codons specifying the rarer transfer RNAs, which would slow down the rate of translation for these genes. The codon distribution of the phage MS2 in table 14.2 shows that every codon is used except the UGA stop codon. (The numbers in the table refer to the incidence of a particular codon in the phage genome.) However, the distribution is not random for all amino acids. For example, the amino acid glycine has two common codons and two rarer codons. The same holds for arginine but not, for example, valine.

Finally, translational control can be exerted at the ribosome. When an uncharged transfer RNA finds its way into

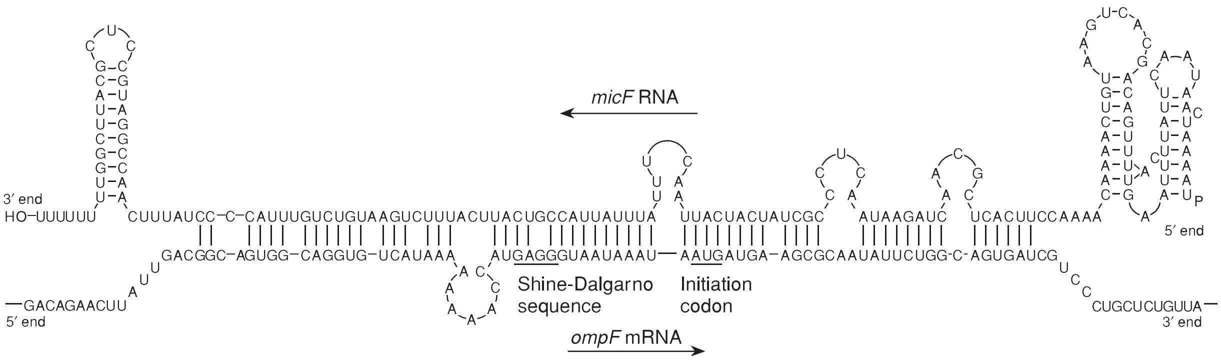
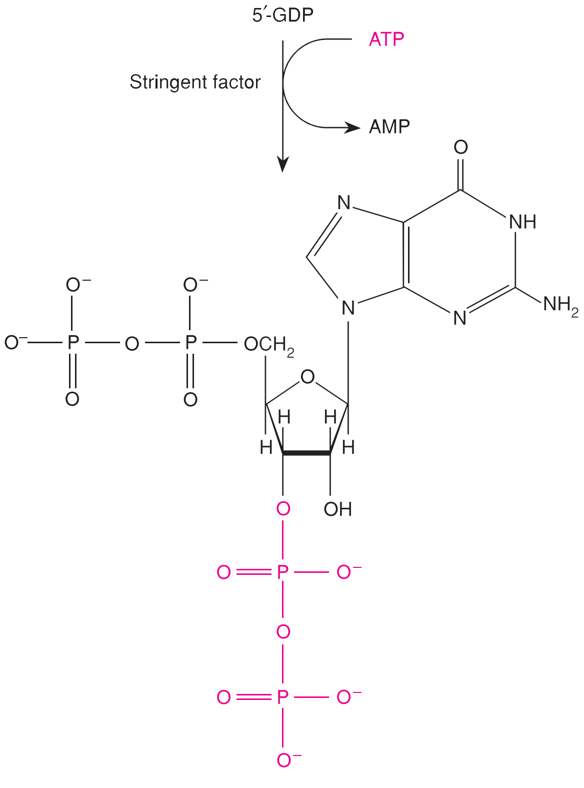


**Figure 14.36** Complementarity between the RNA of the *ompF* gene and antisense RNA, or *micF* RNA. The region of overlap includes the Shine-Dalgarno sequence and the initiation codon, effectively preventing ribosome binding and translation of the *ompF* RNA. Notice the stem-loops on each side of the overlap. (Reproduced, with permission, from the *Annual Review of Biochemistry,* Volume 55, ©1986 by Annual Reviews, Inc.)

the A site of the ribosome, a likely event under amino acid starvation, it causes an **idling reaction** in the ribosome, which entails the production of the nucleotide guanosine tetraphosphate (5′-ppGpp-3′; fig. 14.37). This is part of a control mechanism called the **stringent response.** A protein called the **stringent factor,** the product of the *relA* gene, produces guanosine tetraphosphate (ppGpp), originally called *magic spot* because of its sudden appearance on chromatograms. (The gene is called *rel* from the **relaxed mutant,** which does not have the stringent response.) The stringent factor is associated with the ribosome, where ppGpp is synthesized, although it is not one of the structural proteins of the ribosome. The SpoT protein, the product of the *spoT1* gene, breaks down ppGpp; thus, the concentration of ppGpp is regulated.

The ppGpp interacts with RNA polymerase, causing an almost complete cessation of the transcription of ribosomal RNA; thus, no energy is wasted synthesizing ribosomes when translation is not possible. However, many amino acid-synthesizing operons require ppGpp for transcription; ppGpp thus inhibits ribosomal RNA production and enhances the production of enzymes to synthesize amino acids, all when the cell is starved for amino acids.

One other thing a ribosome can do when faced with amino acid shortages is to slide past *hungry* codons, codons for which a charged transfer RNA is not available. At that point, the growing peptide chain will be attached to the last charged transfer RNA to enter the ribosome, the one now in the peptidyl (P) site. The complex of the transfer RNA and the ribosome slides down the messenger RNA until it encounters the next codon for the transfer RNA. At this point, it is hoped, the next codon en-



ppGpp

**Figure 14.37** The idling reaction. The stringent factor catalyzes the conversion of GDP to 5′-ppGpp-3′. The added pyrophosphate groups come from ATP.

**Table 14.2** Codon Distribution in MS2, an RNA Virus

| | Second Position | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **First Position** | **U** | | **C** | | **A** | | **G** | | **Third Position** |
| **U** | Phe | 10 | Ser | 13 | Tyr | 8 | Cys | 7 | **U** |
| | Phe | 13 | Ser | 10 | Tyr | 13 | Cys | 4 | **C** |
| | Leu | 11 | Ser | 10 | *stop* | 1 | *stop* | 0 | **A** |
| | Leu | 4 | Ser | 13 | *stop* | 1 | Trp | 14 | **G** |
| **C** | Leu | 10 | Pro | 7 | His | 4 | Arg | 13 | **U** |
| | Leu | 14 | Pro | 3 | His | 4 | Arg | 11 | **C** |
| | Leu | 13 | Pro | 6 | Gln | 10 | Arg | 6 | **A** |
| | Leu | 6 | Pro | 5 | Gln | 16 | Arg | 4 | **G** |
| **A** | Ile | 8 | Thr | 14 | Asn | 11 | Ser | 4 | **U** |
| | Ile | 16 | Thr | 10 | Asn | 23 | Ser | 8 | **C** |
| | Ile | 7 | Thr | 8 | Lys | 12 | Arg | 8 | **A** |
| | Met | 15 | Thr | 5 | Lys | 17 | Arg | 6 | **G** |
| **G** | Val | 13 | Ala | 19 | Asp | 18 | Gly | 17 | **U** |
| | Val | 12 | Ala | 12 | Asp | 11 | Gly | 11 | **C** |
| | Val | 11 | Ala | 14 | Glu | 9 | Gly | 4 | **A** |
| | Val | 10 | Ala | 8 | Glu | 14 | Gly | 4 | **G** |

countered in the aminoacyl (A) site will code for a charged transfer RNA present.

## POSTTRANSLATIONAL CONTROL

### Feedback Inhibition

Even after a gene has been transcribed and the messenger RNA translated, a cell can still exert some control over the functioning of the enzymes produced if the enzymes are allosteric proteins. We have discussed the activation and deactivation of operon repressors (e.g., *lac, trp*) owing to their allosteric properties. Similar effects occur with other proteins. The need for posttranslational control is apparent because of the relative longevity of proteins as compared with RNA. When an operon is repressed, it no longer transcribes messenger RNA; however, the messenger RNA that was previously transcribed has been translated into protein, and this protein is still functioning. Thus, during operon repression, it would also be efficient for the cell to control the activity of existing proteins.

An example of posttranslational control occurs with the enzyme aspartate transcarbamylase, which catalyzes the first step in the pathway of pyrimidine biosynthesis in *E. coli* (fig. 14.38). An excess of one of the end products of the pathway, cytidine triphosphate (CTP), inhibits the functioning of aspartate transcarbamylase. This method of control is called **feedback inhibition** because a product of the pathway is the agent that turns the pathway off.

Aspartate transcarbamylase is an allosteric enzyme. Its active site is responsible for the condensation of carbamyl phosphate and L-aspartate (fig. 14.38). However, it also has regulatory sites that have an affinity for CTP. When CTP is bound in a regulatory site, the conformation of the enzyme changes, and the enzyme has a lowered affinity for its normal substrates; recognition of CTP inhibits the condensation reaction the enzyme normally carries out (fig. 14.39). Thus, allosteric enzymes provide a mechanism for control of protein function after the protein has been synthesized.

### Protein Degradation

A final control affecting the amount of gene product in a cell is control of the rate at which proteins degrade. The normal life spans of proteins vary greatly. For example, some proteins last longer than a cell cycle, whereas others may be broken down in minutes. Several models have been suggested for control of protein degradation, including the **N-end rule** and the **PEST hypothesis.**

According to the N-end rule, the amino acid at the amino-, or N-terminal, end of a protein is a signal to proteases that control the average length of life of a protein. In recent experiments, the life span of the β-galactosidase protein was determined with almost complete predictability based on its modified N-terminal amino acid. Protein life spans range from two minutes for those with N-terminal arginine to greater than twenty hours for those with N-terminal methionine or five other amino acids (table 14.3).

According to the PEST hypothesis, protein degradation is determined by regions rich in one of four amino acids: proline, glutamic acid, serine, and threonine. (The one-letter abbreviations of these four amino acids are P, E, S, and T, respectively.) Proteins that have these regions tend to degrade in less than 2 hours. In one study of thirty-five proteins with half-lives of between 20 and 220 hours, only three contained a PEST region. We see that not only are different proteins programmed to survive for varying lengths of time in the cell, but that programming seems to be based on the N-terminal amino acid as well as various regions rich in the PEST amino acids.

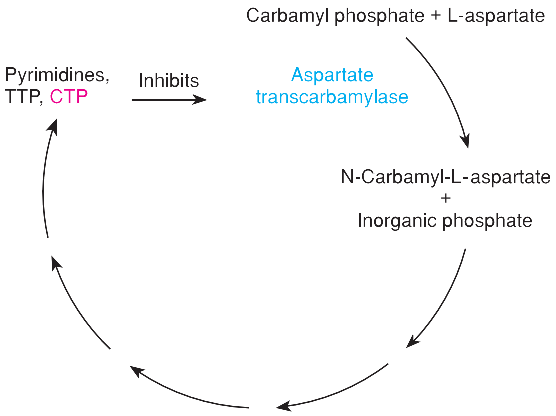

Carbamyl phosphate + L-aspartate
Pyrimidines, TTP, CTP
Inhibits
Aspartate transcarbamylase
N-Carbamyl-L-aspartate + Inorganic phosphate

**Figure 14.38** Aspartate transcarbamylase catalyzes the first step in pyrimidine biosynthesis. An end product, cytidine triphosphate (CTP), inhibits the enzyme.

**Table 14.3 Relationship Between N-Terminal Amino Acid and Half-Life of *E. coli* β-Galactosidase Proteins with Modified N-Terminal Amino Acids**

| N-Terminal Amino Acid | Half-Life |
|---|---|
| Met, Ser, Ala, Thr, Val, Gly | >20 hours |
| Ile, Glu | 30 minutes |
| Tyr, Gln | 10 minutes |
| Pro | 7 minutes |
| Phe, Leu, Asp, Lys | 3 minutes |
| Arg | 2 minutes |

*Source:* Data from Bachmair, et al., *Science,* 234:179–86, 1986.

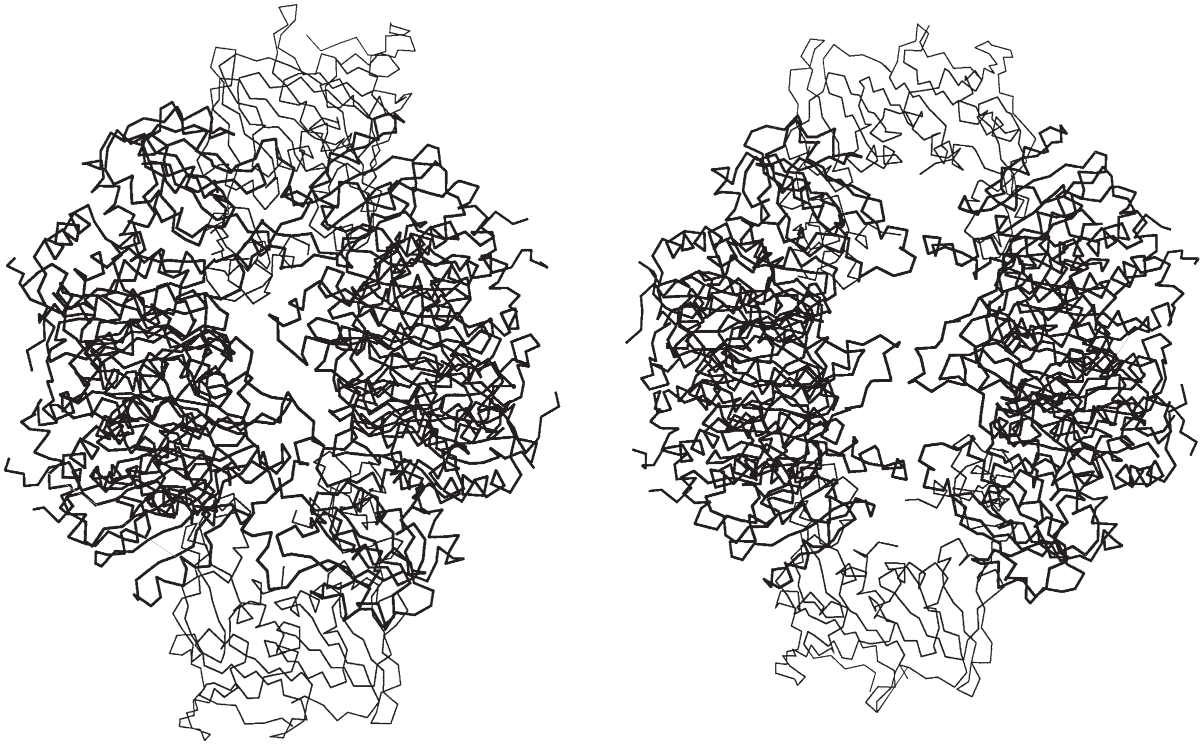


**Figure 14.39** Aspartate transcarbamylase. *Left:* the enzyme bound with cytidine triphosphate (CTP). *Right:* the enzyme without CTP. Notice the difference in shape with and without CTP. With CTP, the enzyme literally closes up. (Courtesy of Evan R. Kantrowitz.)

## SUMMARY

**STUDY OBJECTIVE 1:** To study the way in which inducible and repressible operons work 406–414

Most bacterial genes are organized into operons, which can either be repressed or induced. Transcription begins in inducible operons, such as *lac,* when the metabolite that the operon enzymes act upon appears in the environment. The metabolite (or a derivative), the inducer, combines with the repressor (the product of the independent regulator gene) and renders the repressor nonfunctional. In the absence of the inducer, the repressor binds to the operator, a segment between the promoter and the first gene of the operon. When in place, the repressor blocks transcription. After combining with the inducer, the repressor diffuses from the operator, and transcription proceeds.

All operons responsible for the breakdown of sugars in *E. coli* are inducible. In the presence of glucose, other inducible sugar operons (such as the arabinose and galactose operons) are repressed, even if their sugars appear in the environment. This process is called *catabolite repression.* Cyclic AMP and a catabolite activator protein (CAP) enhance the transcription of the nonglucose sugar operons. Glucose lowers the level of cyclic AMP in the cell and thus prevents the enhancement of transcription of these other operons.

Repressible operons, such as the *trp* operon in *E. coli,* have the same basic traits as an inducible operon—polycistronic transcription controlled by an operator site between the promoter and the first structural gene. However, the repressor protein, controlled by an independent regulator gene, is functional in blocking transcription only after it has combined with the corepressor. This corepressor is the end product of the operon's pathway or some form of the end product (tryptophan in the *trp* operon).

**STUDY OBJECTIVE 2:** To examine attenuator control in bacteria 415–418

Amino acid-synthesizing operons often have an attenuator region. The ability of a ribosome to translate a leader peptide gene determines the secondary structure of the messenger RNA transcript. If the ribosome can translate the leader peptide gene, there must be adequate quantities of the amino acid present, and a terminator stem and loop form in the messenger RNA, causing termination of transcription. In the *trp* operon of bacilli, a protein that binds tryptophans serves the same purpose.

**STUDY OBJECTIVE 3:** To analyze the control of the life cycle of phage λ 418–424

Control of gene expression in λ phage is complex. The "decision" for lytic versus lysogenic response is determined by competition between two repressors, CI and Cro.

**STUDY OBJECTIVE 4:** To determine the way in which transposable genetic elements transpose and control gene expression in bacteria 425–430

Transposons are mobile genetic elements; copies of them can be inserted at other places in the genome. Their ends are inverted repeats. Upon insertion, they are flanked by short, direct repeats. They can be simple (IS elements) or complex. Their presence can cause inversions or deletions. The flagellar phase in *Salmonella* is controlled by the orientation of a transposon.

**STUDY OBJECTIVE 5:** To look at other transcriptional and posttranscriptional mechanisms of control of gene expression in bacteria and phages 430–434

Affinity of early and late operons in phages for different sigma factors are another form of transcriptional control, as seen in phage T4 transcription and the transcription of heat shock proteins. Translational control can be exercised through a gene's position in an operon (genes at the beginning are transcribed most frequently), through redundancy of the genetic code, or through a stringent response that shuts down most transcription during starvation. Posttranslational control is primarily regulated by feedback inhibition. The N-terminal amino acid or particular regions within the proteins program the rate of protein degradation.

## SOLVED PROBLEMS

**PROBLEM 1:** How could you determine whether the genes for the breakdown of the sugar arabinose are under inducible control in *E. coli*?

***Answer:*** *Inducible* means that the genes to break down the substrate—arabinose, a five-carbon sugar, in this case—are not active in the absence of the inducer (again, arabinose). Therefore, in the absence of arabinose in the cells' environment, the arabinose utilization enzymes should not be active within the bacterial cells, but after arabinose is added to the medium, the enzymes should be present. We thus need to assay the contents of the cells before and after arabinose is added to the medium, performing the assay after the cells are broken open and the DNA destroyed so as not to confound the experiment. Using a standard biochemical analysis for arabinose, we should find that the bacterial cell is incapable of metabolizing arabinose before induction but capable of metabolizing it afterward. If the cells were capable of metabolizing arabinose in both cases, we would say that arabinose utilization is constitutive. If the cells were incapable of utilizing arabinose in both cases, we would conclude that the bacterium is incapable of using the sugar arabinose as an energy source. (In fact, arabinose utilization is inducible.)

**PROBLEM 2:** Why would the RecA protein of *E. coli* cleave the λ repressor?

***Answer:*** Since the cleaving of the λ repressor is a signal to begin the lytic phase of the life cycle of the phage, it seems odd that the lysogenized bacterial cell would be an accomplice to its own destruction. However, the phenomenon makes much more sense if we realize that the RecA protein has several other functions critically important to the bacterial cell (see chapter 12). The λ phage has evolved the ability to take advantage of the existence of the RecA protein by evolving a repressor sensitive to it. Evolutionary biologists view this as "coevolution," two interacting organisms evolving to take advantage of or minimize properties of the other. The bacterium, however, might be at a disadvantage. Since RecA has many functions involving interactions with other proteins, it may be highly limited in how it can change. This is one plausible explanation as to why RecA liberates phage λ.

**PROBLEM 3:** What are the differences in action of the λ promoters $p_{RE}$ and $p_{RM}$?

***Answer:*** The promoters $p_{RE}$ and $p_{RM}$ are both promoters of the repressor operon of phage λ. Transcription from these promoters allows production of the ***cI*** repressor protein, the repressor that favors lysogeny. Initially, the promoter $p_{RE}$ is activated. For it to be a transcription site, it must be activated by the product of the ***cII*** gene, which lies in the right operon. This promoter, $p_{RE}$, produces a messenger RNA with a long leader that is translated very efficiently. Once the repressor binds at the operators of the left and right operons, the ***cII*** gene is no longer transcribed, and therefore $p_{RE}$ is no longer a site for transcription. However, the repressor gene can still be transcribed from the $p_{RM}$ promoter, which does not need the product of the *cII* gene. This promoter produces a transcript with no leader and thus is translated very inefficiently. At that point, however, only a very small quantity of repressor is needed to maintain lysogeny. Thus, the two promoters are the sites for the initiation of the repressor operon under different circumstances: one early in the infection stage and one after lysogeny is under way.

**PROBLEM 4:** What are the phenotypes of the following partial diploids for the *lac* operon in *E. coli* in the presence and absence of lactose?

**a.** (F′) $i^+ \; o^+ \; p^+ \; z^-/i^- \; o^+ \; p^+ \; z^+$ (chromosome)
**b.** (F′) $i^+ \; o^c \; p^+ \; z^+/i^+ \; o^+ \; p^- \; z^+$ (chromosome)

***Answer:*** Consider one DNA molecule at a time. If one DNA molecule can never make the enzyme, it can be ignored. In (*a*), the plasmid DNA (F′) will never make enzyme (it is $z^-$), and the chromosomal DNA will never make repressor (it is $i^-$). The functional repressor in the plasmid ($i^+$) will bind to both DNAs, and hence the chromosomal operon will not be transcribed in the absence of lactose and will be induced to transcribe in the presence of lactose. In (*b*), the plasmid DNA (F′) will always be transcribing (operator constitutive) because the repressor can never bind the operator ($o^c$); hence, the operon will be transcribed all the time. The chromosomal DNA can never make RNA (it is $p^-$).

## EXERCISES AND PROBLEMS*

### *LAC* OPERON (INDUCIBLE SYSTEM)

**1.** Are the following *E. coli* cells constitutive or inducible for the *z* gene?
   **a.** $i^+ \; o^+ \; z^+$
   **b.** $i^- \; o^+ \; z^+$
   **c.** $i^- \; o^c \; z^+$
   **d.** $i^+ \; o^c \; z^+$
   **e.** $i^s \; o^+ \; z^+$
   **f.** $i^Q \; o^+ \; z^+$

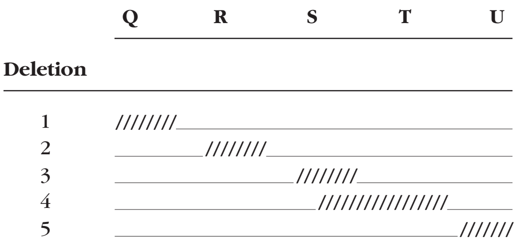
**2.** Determine whether the following *lac* operon merozygotes are inducible or constitutive for the *z* gene.
   **a.** $i^+ \; o^+ \; z^+/\text{F}' \; i^+ \; o^+ \; z^+$
   **b.** $i^- \; o^+ \; z^+/\text{F}' \; i^+ \; o^+ \; z^-$
   **c.** $i^+ \; o^+ \; z^+/\text{F}' \; i^- \; o^+ \; z^-$
   **d.** $i^- \; o^c \; z^-/\text{F}' \; i^+ \; o^+ \; z^+$
   **e.** $i^- \; o^+ \; z^-/\text{F}' \; i^+ \; o^c \; z^+$

**3.** You have isolated a repressor for an inducible operon and have determined that it has two different binding sites, one for the inducer and one for the operator. Mutants of the repressor result in three different phenotypes as far as binding is concerned. What are these phenotypes?

**4.** An *E. coli* strain is isolated that produces β-galactosidase (*lac z*) and permease (*lac y*) constitutively. Provide two possible mutations that could cause this phenotype, and then describe how each mutation would behave in a partial diploid in which the second operon is wild-type for the entire *lac* system.

**5.** You have isolated two *E. coli* mutants that synthesize β-galactosidase constitutively.
   **a.** If these mutants affect different functions, in what two functions could they be defective?
   **b.** You can make a partial diploid of the mutants with the wild-type. What result do you expect for each mutant?

**6.** A hypothetical operon has a sequence of sites, Q R S T U, in the promoter region, but the exact location of the operator and promoter consensus sequences have not been identified. Various deletions of this operator region are isolated and mapped. Their locations appear as follows, with a "/" representing a deleted region.



Deletions 3 and 4 are found to produce constitutive levels of RNA of the operon, and deletion 1 is found to never make RNA. Where are the operator and promoter consensus sequences probably located?

### CATABOLITE REPRESSION

**7.** Describe the role of cyclic AMP in transcriptional control in *E. coli.*

**8.** Operon systems exert negative control by acting through inhibition. The CAP system exerts positive control because it acts through enhancement of transcription. Describe how an operon could work if it were dependent only upon positive control.

---

* Answers to selected exercises and problems are on page A-16.

**9.** J. Beckwith isolated point mutations that were simultaneously uninducible for the *lac, ara, mal,* and *gal* operons, even in the absence of glucose. Provide two different functions that could be missing in these mutants.

### *TRP* OPERON (REPRESSIBLE SYSTEM)

**10.** Construct a merozygote of the *trp* operon in *E. coli* with two forms of the first gene (*e* gene: $e_1$, $e_2$) in the operon. Describe the types of *cis* and *trans* effects that are possible, given mutants of any component of the operon. Can this repressible system work for any type of operon other than those that control amino acid synthesis?

**11.** The tryptophan operon is under negative control; it is on (transcribing) in the presence of low levels of tryptophan and off in the presence of excess tryptophan. The symbols *a, b,* and *c* represent the gene for tryptophan synthetase, the operator region, and the repressor—but not necessarily in that order. From the following data, in which superscripts denote wild-type or defective, determine which letter is the gene, the repressor, and the operator (+ is tryptophan synthetase activity; − is no activity).

| Strain | Genotype | Tryptophan Absent | Tryptophan Present |
|---|---|---|---|
| 1 | $a^- \; b^+ \; c^+$ | + | + |
| 2 | $a^+ \; b^+ \; c^-$ | + | + |
| 3 | $a^+ \; b^- \; c^+$ | − | − |
| 4 | $a^+ \; b^- \; c^+/a^- \; b^+ \; c^-$ | + | + |
| 5 | $a^+ \; b^+ \; c^+/a^- \; b^- \; c^-$ | + | − |
| 6 | $a^+ \; b^+ \; c^-/a^- \; b^- \; c^+$ | + | − |
| 7 | $a^- \; b^+ \; c^+/a^+ \; b^- \; c^-$ | + | + |

**12.** The histidine operon is a repressible operon. The corepressor is charged $\text{tRNA}^{\text{His}}$, and its gene is not part of the operon. For the following mutants, tell whether the enzymes of the operon will be made; then tell whether each mutant would be *cis*-dominant in a partial diploid.

- **a.** RNA polymerase cannot bind the promoter.
- **b.** The repressor-corepressor complex cannot bind operator DNA (the operator has the normal sequence).
- **c.** The repressor cannot bind charged $\text{tRNA}^{\text{His}}$.

### *TRP* OPERON (ATTENUATOR-CONTROLLED SYSTEM)

**13.** Describe the interaction of the attenuator and the operator control mechanisms in the *trp* operon of *E. coli* under varying concentrations of tryptophan in the cell. How does attenuator control react to shortages of other amino acids?

### LYTIC AND LYSOGENIC CYCLES IN PHAGE λ

**14.** What is the fate of a λ phage entering an *E. coli* cell that contains quantities of λ repressor? What is the fate of the same phage entering an *E. coli* cell that contains quantities of the *cro*-gene product?

**15.** Describe the fate of λ phages during the infection process with mutants in the following genes: *cI, cII, cIII, N, cro, att, Q.*

**16.** What is the fate of λ phages during the infection process with mutants in the following areas: $o_{R1}$, $o_{R3}$, $p_L$, $p_{RE}$, $p_{RM}$, $p_R$, $t_{L1}$, $t_{R1}$, *nutL, nutR?*

**17.** What are the three different physical forms that the phage λ chromosome can take?

**18.** How does ultraviolet light (UV) damage induce the lytic life cycle in phage λ?

**19.** The λ prophage is sometimes induced into the lytic life cycle when an Hfr lysogen (lysogenic cell) conjugates with a nonlysogenic $F^-$ cell. How might induction come about in this instance?

**20.** A temperature-sensitive mutant of the λ *cI* gene has been isolated. At 30° C the *cI* repressor binds λ DNA, but it cannot bind DNA at 42° C (it denatures). What is the consequence of incubating *E. coli* that are lysogenic for this λ mutant at 42° C?

**21.** The mutant in problem 20 is heated to 42° C for five minutes, cooled to 30° C, and grown for one hour so that the cells divide several times. The temperature is then raised to 42° C, and you wait for lysis. Many of the cells are not lysed and are in fact able to form colonies. Explain these results.

### TRANSPOSABLE GENETIC ELEMENTS

**22.** Why are IS elements sometimes referred to as "selfish DNA"?

**23.** What are the differences among an IS element, a transposon, an intron, a plasmid, and a cointegrate?

**24.** Describe the Shapiro model of transposition. What are the roles of transposase, DNA polymerase I, ligase, and resolvase?

**25.** Why are transposons flanked by direct repeats?

**26.** How do transposons induce deletions? inversions?

**27.** Describe how a transposon controls the expression of the flagellar phase in *Salmonella.*

**28.** What is a polar mutation? What can cause it?

### OTHER TRANSCRIPTIONAL CONTROL SYSTEMS

**29.** List the steps from transcription through translation to enzyme function, noting all the points at which control could be exerted. (*See also* TRANSLATIONAL CONTROL and POSTTRANSLATIONAL CONTROL)

**30.** What are the advantages of transcriptional control over translational control? (*See also* TRANSLATIONAL CONTROL)

**31.** How are heat shock proteins induced?

**32.** In phage T4, the genes *rIIA* and *rIIB* lie adjacent to each other on the T4 chromosome. During the early phase of infection, *rIIA* and *rIIB* products are present in equimolar amounts. In the late phase of infection, the amount of *rIIB* protein is ten to fifteen times higher than that of *rIIA* protein. Nonsense mutations (mutations to a stop codon) in *rIIA* eliminate early but not late *rIIB* transcription. In the mutants that contain small deletions near the end of *rIIA,* the amount of *rIIA* product is always equal to the amount of *rIIB* product, regardless of the time of infection. Based on this information, devise a map of the *rII* region. Include the location(s) of the promoter(s).

**TRANSLATIONAL CONTROL**

**33.** What is the stringent response? How does it work? What is an idling reaction?

**34.** What is antisense RNA? How does it work? What is the obvious source of this regulatory RNA? How could this RNA be used to treat a disease clinically?

**POSTTRANSLATIONAL CONTROL**

**35.** What is feedback inhibition? What other roles do allosteric proteins play in regulating gene expression?

**36.** What controls the rate of degradation of proteins?

## CRITICAL THINKING QUESTIONS

**1.** From an evolutionary perspective, why do you think *Escherichia coli* evolved a CAP system of positive control of gene expression? Why not just metabolize any and all sugars in the environment as they appear?

**2.** Why might some proteins and messenger RNAs produced in *Escherichia coli* be degraded so quickly?

*Suggested Readings for chapter 14 are on page B-13.*

# 15

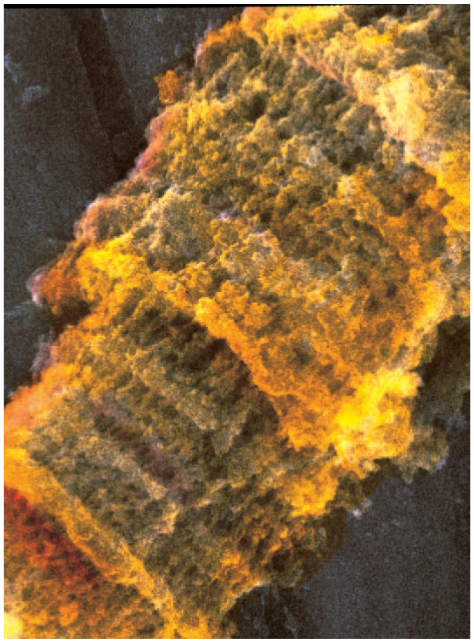
# THE EUKARYOTIC CHROMOSOME

## STUDY OBJECTIVES

1. To examine the arrangement of DNA and proteins comprising the eukaryotic chromosome 440
2. To look at the nature of centromeres and telomeres in eukaryotic chromosomes 453
3. To analyze the nature of the DNA in eukaryotic chromosomes 457

## STUDY OUTLINE

**The Eukaryotic Cell** 440
**The Eukaryotic Chromosome** 440
  DNA Arrangement 440
  Nucleoprotein Composition 442
  Chromosomal Banding 451
  Centromeres and Telomeres 453
  The C-Value Paradox 457
*Summary* 461
*Solved Problems* 462
*Exercises and Problems* 462
*Critical Thinking Questions* 463
**Box 15.1 How Big Is Big, How Small Is Small?** 441
**Box 15.2 High-Speed Chromosomal Sorting** 444



Artificially colored scanning electron micrograph of part of a polytene salivary gland chromosome from a fruit fly (*Drosophila*), revealing the underlying banding pattern. (© Professors P. Motta and T. Naguro/ SPL/Photo Researchers. inc.)

In chapter 14, we looked at the control of gene expression in prokaryotes and bacteriophages. Compared to eukaryotes, bacteriophages and prokaryotes are relatively simple. Of fundamental importance is that, in these lower forms, the operon model of induction and repression of transcription is a unifying theme for control of gene expression. Despite nuances such as catabolite repression and attenuator control, the operon model provides a relatively clear picture of how genes are turned on and off in phages and prokaryotes. This model does not exist for eukaryotes. In attempting to elucidate models for control of gene expression in eukaryotes, we must take one very important factor into account: the complexity of the structure of the eukaryotic chromosome. In this chapter, we cover the current understanding of how these very large structures are organized.

## THE EUKARYOTIC CELL

Eukaryotes and prokaryotes are the two superkingdoms of organisms. The following comparisons, using *E. coli* as a general model for prokaryotes, show how much more complex eukaryotes are:

1. An *E. coli* chromosome contains approximately $4.2 \times 10^6$ base pairs of DNA. The haploid human genome contains nearly one thousand times as much DNA.
2. Eukaryotic DNA is in the form of nucleoprotein, a DNA-histone protein complex. Although a few histonelike proteins have been found in *E. coli,* its chromosomal DNA is not complexed with protein to anywhere near the same extent.
3. An *E. coli* cell has very little internal structure. Eukaryotes have a number of internal organelles and an extensive lipid membrane system, including the nuclear envelope itself.
4. An *E. coli* cell is small (0.5 to 5.0 μm in length for bacteria). Eukaryotic cells are generally larger than prokaryotes (10 to 50 μm in length for animal tissue cells; box 15.1).
5. The messenger RNA of *E. coli* is translated while it is being transcribed. Eukaryotic messenger RNA is modified within the nucleus before it is transported out for translation in the cytoplasm.
6. Almost no messenger RNA isolated from eukaryotic cells, including the messenger RNA of animal viruses, has been found to be polycistronic (containing many genes). Most prokaryotic messenger RNAs are polycistronic.
7. Most *E. coli* genes are parts of inducible or repressible operons; there are almost no operons in eukaryotes.
8. *E. coli* exists as a simple, single cell. Although some prokaryotes do aggregate, sporulate, and show a few other limited forms of differentiation, they are primarily one-celled organisms. And, although some eukaryotes are single-celled (e.g., yeast), the essence of eukaryotes is differentiation. In human beings, a zygote gives rise to every other cell type in the body in a relatively predictable manner.

To fully appreciate the complexity of eukaryotes, we begin by looking at the eukaryotic chromosome. In the next chapter, we look at the patterns of development in eukaryotes and some mechanisms of control of gene expression.

## THE EUKARYOTIC CHROMOSOME

### *DNA Arrangement*

Evidence that the eukaryotic chromosome is **uninemic**—that is, contains one double helix of DNA—comes from several sources. The best data are provided by radioactive-labeling studies, first done by J. Taylor and his colleagues in 1957. If a eukaryote is allowed to undergo one DNA replication in the presence of tritiated ($^3$H-) thymidine, each of the daughter chromatids would be expected to contain a double helix with one unlabeled DNA template strand and one labeled strand of newly synthesized bases (fig. 15.1). This configuration is expected on the basis of semiconservative replication, with each chromatid containing one double helix. A second round of DNA replication, in the absence of $^3$H-thymidine, should produce chromosomes in which one chromatid would have unlabeled DNA and one would have labeled DNA. Figure 15.2 shows the chromosomes after this second replication in nonlabeled media. As expected, one chromatid of every pair is labeled and one is not.

In another kind of experiment, R. Kavenoff, L. Klotz, and B. Zimm demonstrated that *Drosophila* nuclei contained pieces of DNA of the size predicted from their DNA content, based on the premise that each chromosome contains one DNA molecule. They isolated the DNA and measured the size of the largest DNA molecules using the *viscoelastic* property of DNA, the rate at which

Ruth Kavenoff (1944– ). (Courtesy of Dr. Ruth Kavenoff.)

## BOX 15.1

## Experimental Methods

### *How Big Is Big, How Small Is Small?*

Generally, eukaryotic cells are large, and prokaryotic cells are small. For example, an average eukaryotic cell is about 50 μm in diameter, whereas an average bacterium is about 5 μm in length. The average virus is about 0.05 μm in diameter. These size differences occur because eukaryotic cells have complex substructures and internal architecture that prokaryotic cells lack. Since we believe that prokaryotic cells depend on diffusion to exchange materials with the environment, they would have to be small. And viruses, intracellular parasites, would of necessity be very small. There are, of course, exceptions.

In 1999, a team of scientists from Germany, Spain, and the United States isolated large sulfur bacteria off the Namibian coast of Africa and named them *Thiomargarita namibiensis,* the sulfur pearl of Namibia. These bacteria can be almost half a millimeter in diameter, the size of the period at the end of this sentence (fig. 1). Based on the sequence of 16S ribosomal DNA, these bacteria were shown to be close relatives of other marine sulfur bacteria. They are almost one hundred times the volume of the bacteria previously believed to be largest, *Epuliscium fishelsoni,* known only from the intestine of the brown surgeonfish.

The smallest prokaryotes are the Mycoplasmas, at about 0.2 μm in diameter, rivaling the viruses in size. They are animal pathogens and decomposing organisms. The smallest eukaryote, *Ostreococcus tauri,* a green alga found in the plankton, was discovered in 1994 from a water sample in a French lagoon on the Mediterranean Sea by a group of French scientists. These organisms are less than 1 μm in diameter. Scientists believe that the lower limit on the size of a cell (not counting viruses) is about 200 nm (0.2 μm), based on the size of DNA and ribosomes that a cell must contain.

With *T. namibiensis* as the largest prokaryote, we note that the largest eukaryotic cell with a single nucleus is most likely the ostrich egg. The largest organisms are the blue whale, *Balaenoptera musculus,* weighing in at 118,000 kilograms; giant redwood trees, *Sequoiadendron giganteum,* 100 meters tall and weighing 5.5 million kilograms; a quaking aspen clone, *Populus tremuloides,* weighing 6 million kilograms; and *Armillaria bulbosa,* a fungus. In 1992, three scientists from the University of Toronto and Michigan Technological University, using restriction fragment length polymorphisms (RFLPs) and polymerase chain reaction (PCR) techniques, showed that the huge hyphal mass of this tree-root colonizing fungus growing in a forest in northern Michigan was a single organism. It covered about eight hectares, probably weighed more than 10,000 kilograms, and probably has existed for more than 1,500 years.

Although we don't want to get distracted by the oddities and extremes of nature, size differences are remarkable.

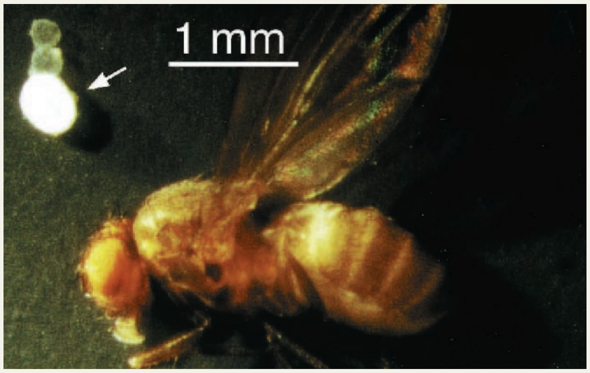


**Figure 1** The bacterium *Thiomargarita namibiensis* shown with a fruit fly (*Drosophila virilis,* 3 mm in length) for size comparison. The arrow points to a single bacterial cell, 0.5 mm wide, bright with sulfur inclusions. Above the cell are empty sheaths of dead bacteria. (From H.N. Schulz, et al., "Dense populations of a giant sulfur bacterium in Namibian Shelf Sediments" in *Science,* Vol. 284, pp. 493–95, April 16, 1999. Reprinted by permission of the American Association for the Advancement of Science.)

stretched molecules relax. From other sources, primarily UV absorbance studies, it was estimated that the largest *Drosophila* chromosome had about $43 \times 10^9$ daltons of DNA. Results from the viscoelastic measurements indicated the presence of DNA molecules of between 38 and $44 \times 10^9$ daltons. Viscoelastic measurements of inversions, which changed the ratio of the arms but not the overall size of the chromosome, yielded similar results. However, a translocation that radically changed the size of the chromosome to $59 \times 10^9$ daltons resulted in an

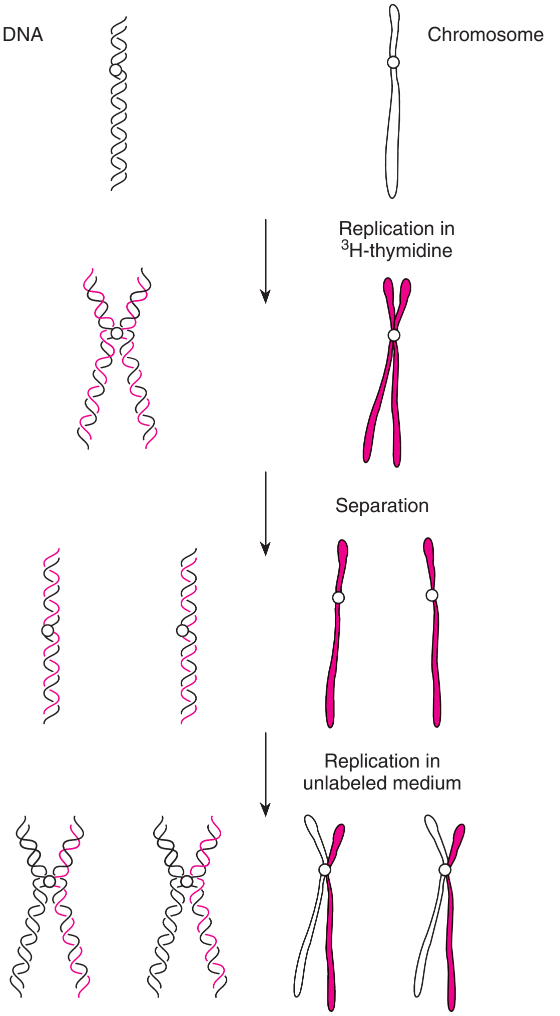

DNA
Chromosome
Replication in $^{3}$H-thymidine
Separation
Replication in unlabeled medium

**Figure 15.1** Radioactive labeling of a uninemic eukaryotic chromosome following semiconservative replication. Replication occurs first in the presence of $^{3}$H-thymidine and then in its absence. *Red* represents labeling. After the second round of replication, one chromatid of each chromosome is labeled, whereas the other is not, confirming that there is only one DNA molecule per chromatid and that the chromosome is thus uninemic.

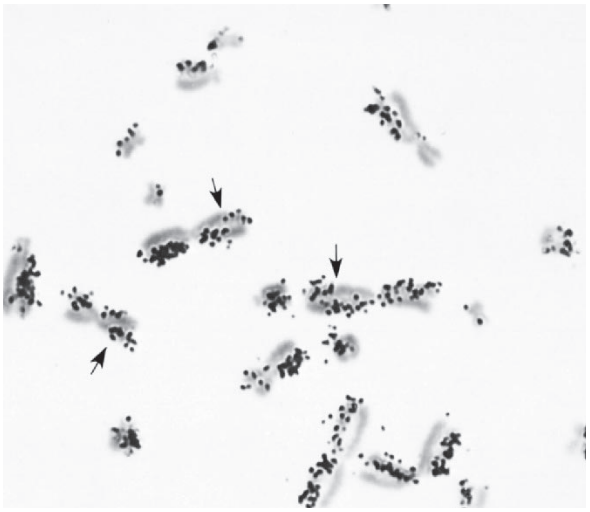


**Figure 15.2** Second metaphase in hamster cells in culture after one replication in the presence of $^{3}$H-thymidine followed by one in nonradioactive medium, verifying the uninemic nature of the eukaryotic chromosome. Cases in which the label apparently switches from one chromatid to the other are caused by sister chromatid exchanges (at *arrows*). (Source: G. Marin and D. M. Prescott, "The frequency of sister chromatid exchanges following exposure to varying doses of $^{3}$H-thymidine or X-ray," *Journal of Cell Biology,* 21, (1964): 159–67, by copyright permission of the Rockefeller University Press.)

equivalent change in the viscoelastic estimates to between 52 and 64 × $10^9$ daltons.

The conclusion from these studies is that the largest *Drosophila* chromosome, and by extension every eukaryotic chromosome, contains a single DNA molecule running from end to end, encompassing both arms. The viscoelastic values were corroborated by carefully isolating and measuring the lengths of long DNA molecules, an especially difficult task given DNA's propensity to break. The longest molecule that the investigators found was 1.2 cm long, equivalent to between 24 and 32 × $10^9$ daltons (fig. 15.3), close to the predicted size. Thus, the evidence is in complete concordance with the simple uninemic model of eukaryotic chromosomal structure (box 15.2).

## *Nucleoprotein Composition*

### *Nucleosome Structure*

Since each eukaryotic chromosome consists of a single, relatively long piece of duplex DNA, the average diploid cell contains many of these long pieces of DNA. For chromosomes to be properly distributed to each daughter cell during mitosis and meiosis, they must be condensed into structures that are more easily managed. Wrapping the DNA around "spools" of protein constitutes the first step in a series of coiling and folding processes that eventually result in the fully compacted chromosome we see at metaphase.

Interphase nuclei can be disrupted by placing them in a hypotonic liquid such as water. When this happens, chromatin material is released. When this material is observed under the electron microscope, small particles

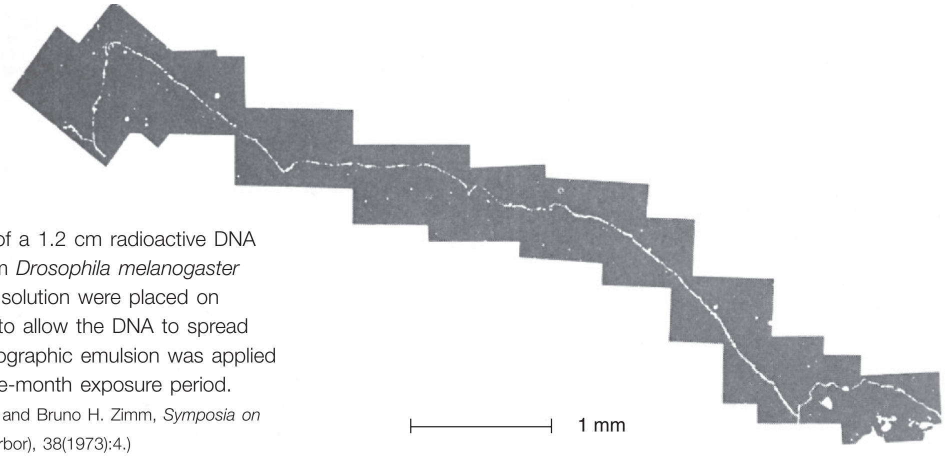


**Figure 15.3** Autoradiograph of a 1.2 cm radioactive DNA molecule carefully isolated from *Drosophila melanogaster* chromosomes. Drops of DNA solution were placed on microscope slides, then tilted to allow the DNA to spread slowly down the slide. A photographic emulsion was applied and later developed after a five-month exposure period. (From Ruth Kavenoff, Lynn C. Klotz, and Bruno H. Zimm, *Symposia on Quantitative Biology* (Cold Spring Harbor), 38(1973):4.)

called **nucleosomes** can be seen (fig. 15.4). These are the spools that the DNA is wrapped around. They are made of **histone** proteins and associated DNA (table 15.1). The histones, a group of arginine- and lysine-rich basic proteins, have been well characterized. They are especially well suited to bind to the negatively charged DNA (table 15.2).

When chromatin is treated with micrococcal nuclease, individual nucleosomes can be isolated, indicating that the DNA between nucleosomes is accessible to digestion. The results of these studies indicate that a length of 168 base pairs (bp) of DNA, the core DNA, is intimately associated with the nucleosome, and another 50 to 75 base pairs, depending on species, connects the nucleosomes (linker DNA; fig. 15.5). When the quantities of the various histones were measured, there were two each of histones H2A, H2B, H3, and H4 per nucleosome and only one molecule of histone H1. Reconstitution and degradation studies have indicated that histone H1 is not a necessary component in the formation of nucleosomes. We believe that histone H1 is associated with the linker DNA as it enters and emerges from the nucleosome (fig. 15.6), although its exact position is not known with certainty. Histone H1 may be more off center and internally located than illustrated. The term **chromatosome** has been suggested for the core nucleosome plus the H1 protein, a unit that includes approximately 168 base pairs of DNA. Nucleosomes, then, are a first-order packaging of DNA; they reduce its length and undoubtedly make the coiling and contraction required during mitosis and meiosis more efficient (fig. 15.7).

When DNA is replicated, twice as many nucleosomes are needed since one double helix becomes two. Recent studies indicate that a parental nucleosome is partly disassembled during DNA replication and reassembled on one or the other daughter strand, apparently randomly. The other DNA strand has a new nucleosome constructed

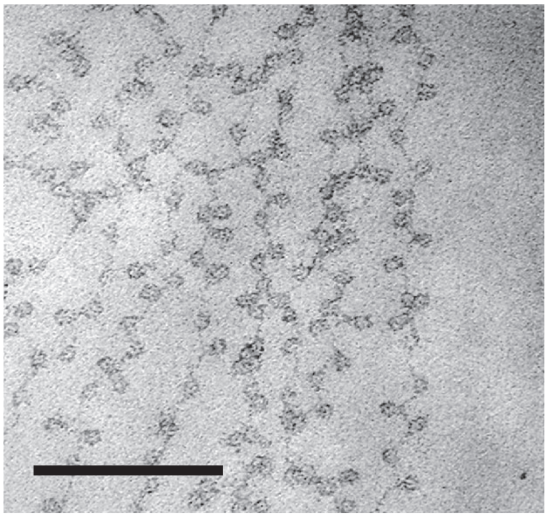


**Figure 15.4** Electron micrograph of chromatin fibers. Photo shows nucleosome structures (*spheres*) and connecting strands of DNA called linkers. The bar is 100 nm long. (Source: D. E. Olins and A. L. Olins, "Nucleosomes: The structural quantum in chromosomes," *American Scientist,* 66: 704–11, November 1978. Reproduced by permission.)

**Table 15.1** **The Constituency of Calf Thymus Chromatin**

| Constituent | Relative Weight* |
|---|---|
| DNA | 100 |
| Histone proteins | 114 |
| Nonhistone proteins | 33 |
| RNA | 1 |

* Weight relative to 100 units of DNA.

## BOX 15.2

## Experimental Methods

### *High-Speed Chromosomal Sorting*

To facilitate the creation of recombinant genomic libraries, for mapping purposes, and for other reasons, it is useful to be able to isolate individual human chromosomes. To these ends, several methods have been developed to isolate chromosomes. Here we discuss a high-speed sorting method based on fluorescent staining and flow cytometry.

DNA can be treated with several fluorescent dyes. Chromosomes can then be recognized individually by their relative fluorescent intensities. The dyes Hoechst 33258 and chromomycin A3 are a valuable combination because they respond to different wavelengths of light and they bind DNA differently. Hoechst binds preferentially to DNA rich in adenine and thymine, whereas chromomycin binds preferentially to DNA rich in guanine and cytosine. Thus, since every human chromosome has a unique ratio of bases, the relative intensity of each chromosome is different when fluoresced.

Chromomycin fluoresces in the presence of a laser tuned to 458 nm, and Hoechst fluoresces in the presence of a UV laser. The chromosomes can be identified when their relative fluorescence in the two lasers is plotted, producing a *flow karyotype* (fig. 1). Modern flow cytometry techniques then allow the isolation of these identified chromosomes.

In practice, chromosomes are isolated in large numbers from cells that have been arrested in metaphase by treatment with colcemid, which inhibits spindle formation. These chromosomes are then purified in buffer and treated with the two dyes. The chromosomes are separated at high speed (two hundred chromosomes per second) in a flow cytometry device (fig. 2). As the chromosome-containing buffer passes through the laser beams, identification is made. The liquid is then forced to form minute droplets (215,000 per second) by passing through a vibrator. Specific droplets carrying the identified chromosomes are then charged, either positively or negatively, and passed between deflection plates. Positively charged droplets pass one way, and negatively charged droplets pass the other way, thus allowing the simultaneous isolation of two different chromosomes. At a rate of two hundred chromosomes per second, it

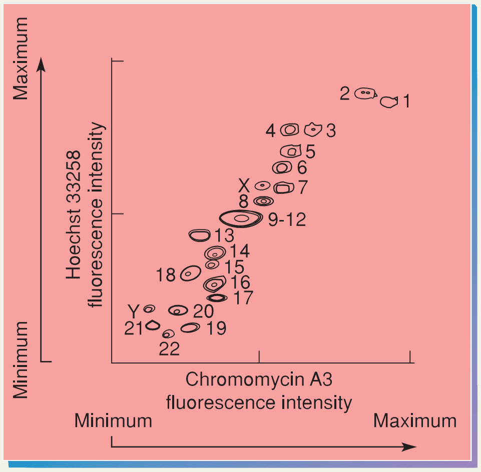


**Figure 1** Flow karyotype of human chromosomes at very high resolution, measured under low-speed sorting (fifteen to thirty-five chromosomes per second). The ordinate is Hoechst 33258 fluorescence intensity, and the abscissa is chromomycin A3 fluorescence intensity. All chromosomes are resolved except numbers 9–12. (Reprinted with permission from J. W. Gray, et al., "High-Speed Chromosome Sorting," *Science,* 238:323–329, 1987. Copyright © 1987 American Association for the Advancement of Science.)

of histones from the cellular pool with the help of proteins called **chromatin assembly factors;** at least three of these factors are known.

For example, in fruit flies, a protein complex called the **replication-coupling assembly factor** assembles new nucleosomes. In addition, a protein complex called **condensin** is needed for the condensation of interphase chromosomes to mitotic chromosomes. This complex includes two **SMC proteins** (for *s*tructural *m*aintenance of *c*hromosomes) and two non-SMC proteins. SMC proteins also aid other chromosomal activities, such as mitotic segregation, sister-chromatid adhesion, dosage compensation, and recombination. Thus, a diverse array of proteins is involved in creating nucleosomes and chromatois possible to isolate 0.1 g of DNA in less than an hour; 0.1 g of DNA is adequate for library construction and represents about $5 \times 10^5$ average chromosomes.

The technique is not perfect. During isolation, debris and clumps of chromosomes are produced that cause contamination problems. Then, some chromosomes are so similar in their fluorescence that they are hard to separate. This is true, for example, for chromosomes 9 to 12. Also, chromosome 21 is hard to separate because its fluorescence tends to fall into the debris area.

Some of these problems, however, can be overcome by using hybrid cell lines of hamsters, for example, containing only one human chromosome. It is much easier to isolate the human chromosome from the hybrid line. Purity values of 90% are not unreasonable, with some in excess of 95%.

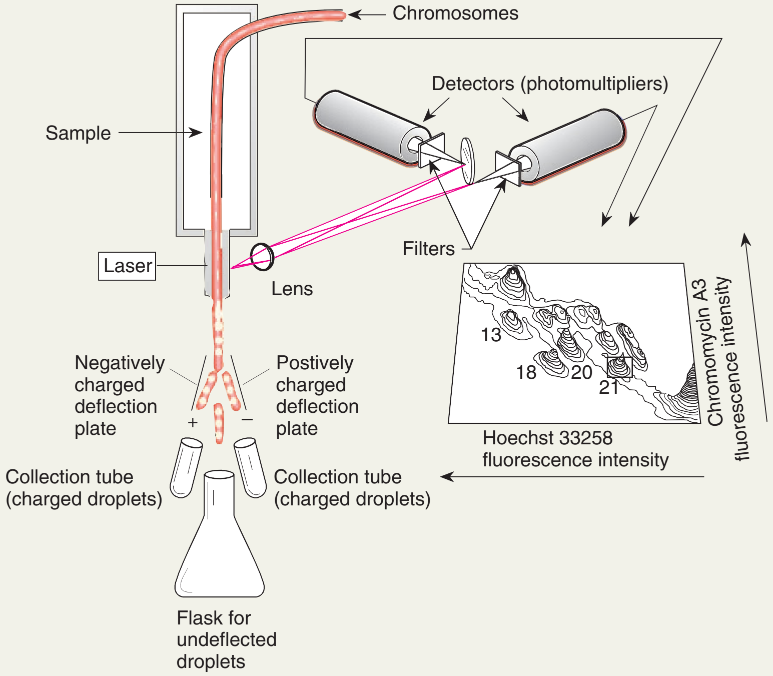

Chromosomes
Detectors (photomultipliers)
Sample
Filters
Laser
Lens
13
18
20
21
Chromomycin A3 fluorescence intensity
Hoechst 33258 fluorescence intensity
Negatively charged deflection plate
Postively charged deflection plate
+
−
Collection tube (charged droplets)
Collection tube (charged droplets)
Flask for undeflected droplets

**Figure 2** The flow cytometry device used to separate chromosomes at high speed. A buffer with chromosomes enters the device. Lasers cause fluorescence that is analyzed with the aid of the photomultiplier tubes. Droplet formation is induced by vibration, and, based on a flow rate of 50 m/sec, appropriate drops are charged. Charged drops are then separated by charged deflection plates and collected. Uncharged droplets pass through. (Reprinted with permission from J. W. Gray, et al., "High-Speed Chromosome Sorting," *Science,* 238:323–329, 1987. Copyright © 1987 American Association for the Advancement of Science.)

somes, condensing interphase chromosomes, and performing numerous other activities of chromosomes.

Dosage compensation has recently been associated with a change in nucleosome structure. The inactivated X chromosome appears to have a different type of histone present. Histone H2A is replaced by a variant called mH2A. The details of this mechanism are under study.

Nucleosomes apparently play a major role in controlling gene expression; DNA with nucleosomes has a much lower transcription rate than DNA without nucleosomes. It makes sense that the positions of nucleosomes can provide or prevent access to promoters. There are regions of the DNA, known as **nuclease-hypersensitive sites,** that appear to be nucleosome free. These sites, usually mutiples of a nucleosomal region of about two hundred base pairs, are particularly sensitive to digestion by different nucleases. When these regions are isolated, they usually have sequences that control functions in replication, transcription, or other activities of DNA. For example, numerous promoter regions in *Drosophila,* mouse, and human DNA are in nuclease-hypersensitive sites. Hence, some specific DNA sequences are kept free of nucleosomes, and these sequences appear to be recognized by various enzymes such as RNA polymerase. In many other cases, however, nucleosomes do appear to cover promoters and repress transcription. For transcription to occur in these cases, some form of **chromatin remodeling** must take place.

Two general classes of proteins are involved in chromatin remodeling. First are proteins that acetylate the N-terminal tails of the histones, a process that may cause the nucleosomes to bind the DNA less tightly and thus make it available for attachment of transcription factors. These enzymes are called **histone acetyl transferases (HATs).** Deacetylating enzymes have the reverse effect: They act to repress transcription. Second, a class of ATP-dependent proteins such as the SWI/SNF complex in yeast also affect chromatin remodeling. (Some workers called the proteins SWI because they were involved in mating type *swi*tching, and others called them SNF for *s*ucrose *n*on*f*ermenting.) The SWI/SNF complex is a group of eleven proteins involved in transcription activation in many genes, presumably allowing transcription

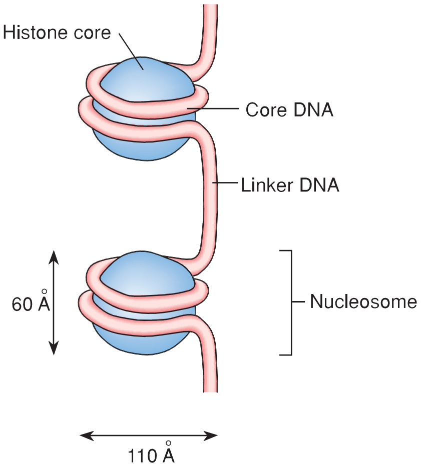


**Figure 15.5** The eukaryotic chromosome is associated with histone proteins to form nucleosomes. The protein core is wrapped with 1.7 loops of DNA and connected with a length of DNA called a linker.

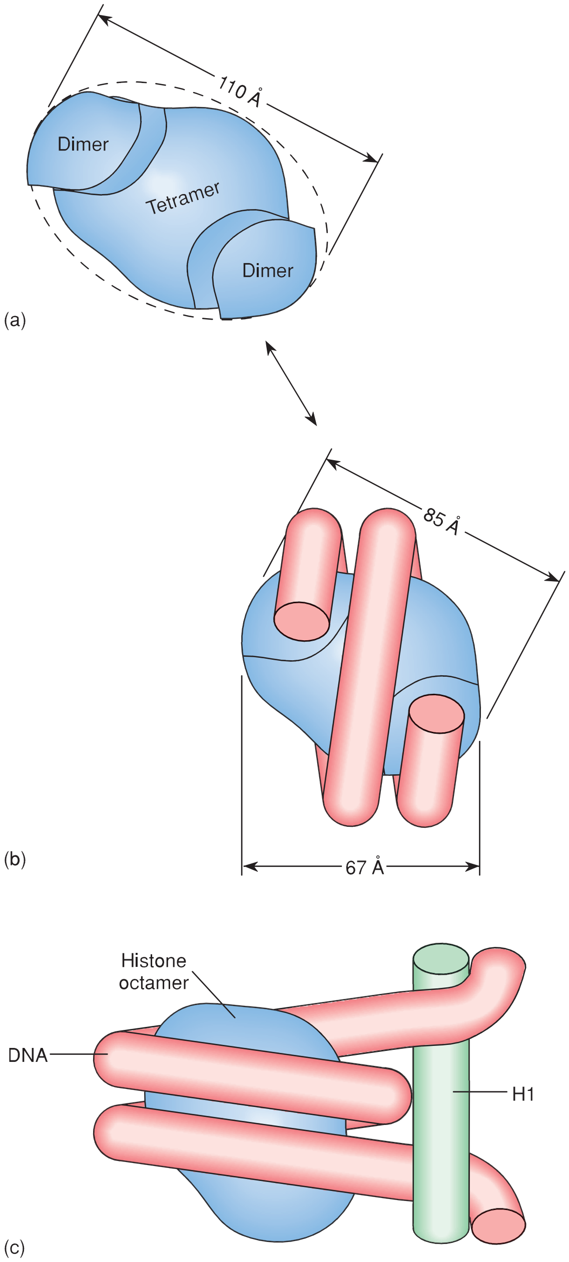


**Figure 15.6** Nucleosome structure. (*a*) Schematic comparison of the eight histones comprising the nucleosome in salt solution. A dimer consists of one H2A and one H2B histone molecule; a tetramer consists of two H3 and two H4 histones. (*b*) DNA fits in surface grooves on the more compacted structure found in physiological conditions. (*c*) The diagram shows the presumed position of the H1 histone, encompassing 166 base pairs of DNA.

**Table 15.2** Composition of Histones

| Fraction | Class | Number of Amino Acids | Percentage of Basic Amino Acids |
|---|---|---|---|
| H1 | Very lysine rich | 213 | 30 |
| H2A | Lysine, arginine rich | 129 | 23 |
| H2B | Moderately lysine rich | 125 | 24 |
| H3 | Arginine rich | 135 | 24 |
| H4 | Arginine, glycine rich | 102 | 27 |

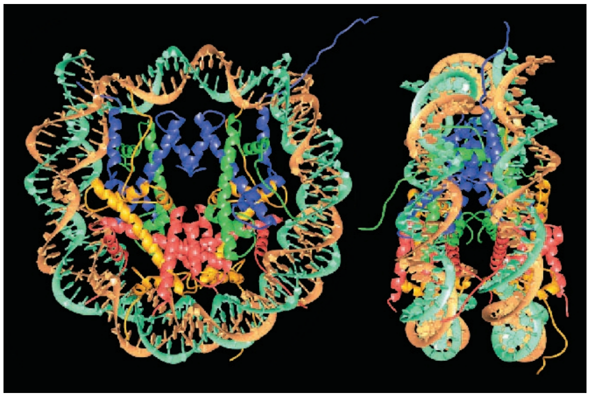


**Figure 15.7** Nucleosome core particle at 2.8 Å resolution. Shown are 146 base pairs of DNA (*brown* and *turquoise*) and the eight histone protein chains (*purple:* H3; *green:* H4; *yellow:* H2A; and *red:* H2B). Note the protein tails (N-terminal ends) of the histone polypeptides extending out of the nucleosome. On the *left* is the view down the DNA helix, and on the *right* is the perpendicular view. (From Karolin Luger, et al., "Crystal structure of the nucleosome core particle at 2.8A resolution" in *Nature,* 389:251–260, September 18, 1997, fig. 1a p. 252. Reprinted by permission of Macmillan, Ltd.)

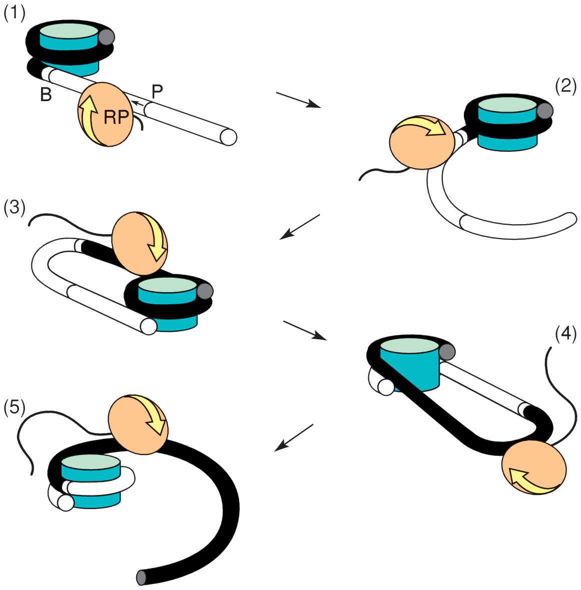


**Figure 15.8** RNA polymerase steps around a nucleosome without disrupting it. (*1*) The RNA polymerase begins at a promoter (*P*) and heads for the nucleosomed DNA (filled in), whose border is noted with a line and the letter *B.* As the polymerase encounters the nucleosome, it begins to unwrap the DNA from the histones (*2*). The displaced DNA then reencounters the histones (*3*), about seventy-five to eighty base pairs upstream from the original point of nucleosome formation. The polymerase continues on its way (*4* and *5*), and the nucleosome re-forms in its displaced position without disrupting the histones or ever fully losing contact with the core DNA. (From Vasily M. Studitsky, et al., "A histone octamer can step around a transcribing polymerase without leaving the template," *Cell,* 76: 371–82, January 28, 1994. Copyright © 1994 by Cell Press. Reprinted by permission.)

factors to access promoters by remodeling chromatin. These proteins are able to reposition a nucleosome on DNA by sliding the nucleosome down the DNA.

We thus conclude that although nucleosomes serve as a general, first-order packing mechanism in eukaryotic DNA, they can be positioned precisely and can attenuate transcription. It is interesting to note that once transcription begins, RNA polymerase apparently moves along nucleosomed DNA by translocation of the histones by 75 to 80 base pairs without disrupting the nucleosome itself. This seems to be accomplished by the RNA polymerase moving the DNA and then re-forming the nucleosome in its wake (fig. 15.8).

### *Higher-Order Structure of Chromatin*

Since the nucleosome has a width of only 110 Å, and metaphase chromosomes appear to be constructed of a fiber having a diameter of about 2,400 Å (fig. 15.9), several additional levels of chromatin compaction lead to the metaphase chromosome. Various experiments, which change the ionic strength the chromatin is subjected to, indicate that the 110 Å DNA spontaneously forms a 300 Å, solenoidlike fiber with increased ionic strength. It seems that this fiber results from the coiling of the nucleosomal DNA (fig. 15.10). This 300 Å fiber is not, however, the final form of the DNA. We can account for the contraction of the 300 Å fiber to the 2,400 Å fiber found in metaphase chromosomes by the formation of a second solenoidlike structure from the winding of the 300 Å fiber (fig. 15.11).

If the histones are removed from a chromosome, the DNA billows out, leaving a proteinaceous structure termed a **scaffold** (fig. 15.12). This scaffold structure is formed from **nonhistone proteins;** two of them predominate, namely SC1 and SC2. SC1 has been identified as topoisomerase II. It would not be unreasonable to expect several hundred different proteins, many in minute quantities, to be associated with the chromosome and involved in replication, repair, and transcription.

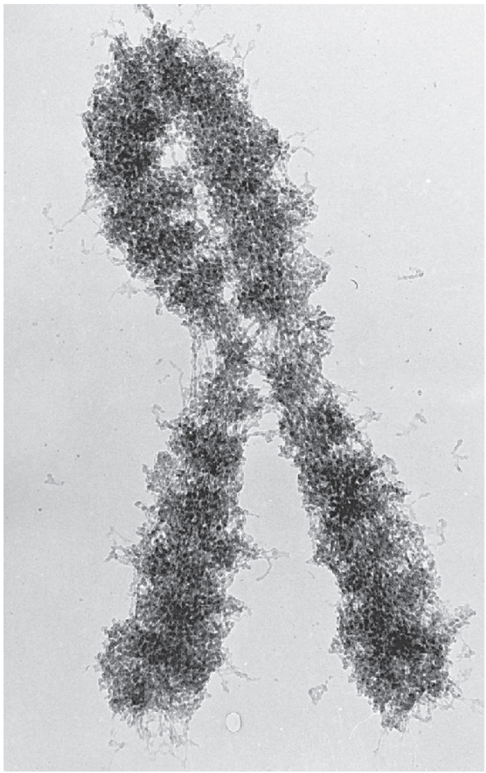


**Figure 15.9** Chinese hamster chromosome. Note the fibers making up the chromosome; they are approximately 2,400 Å in diameter. Magnification 11,800×. (Source: Courtesy of Dr. Hans Ris.)

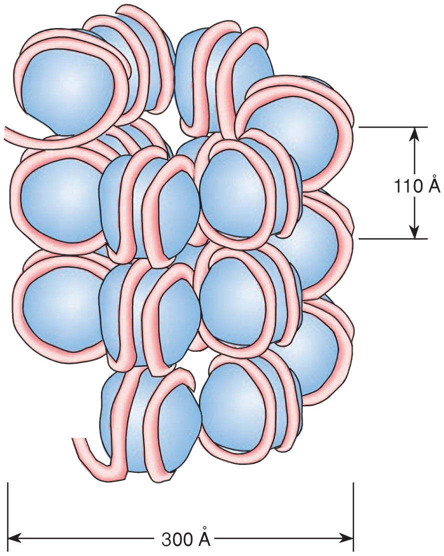


**Figure 15.10** Solenoid model for the formation of the 300 Å chromatin fiber. Nucleosomal DNA wraps in a helical fashion, forming a hollow core. Although histone H1 is not shown, it is known to be on the inside of the solenoid.

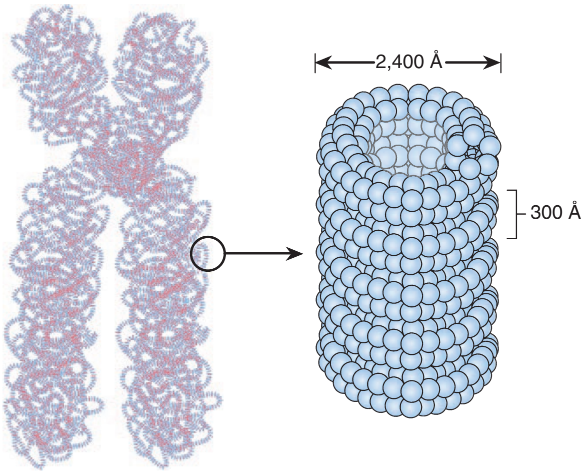


**Figure 15.11** The 2,400 Å fiber of the eukaryotic chromosome is a hollow, solenoidlike structure. It is formed by the coiling of the 300 Å fiber, which itself is a solenoid.

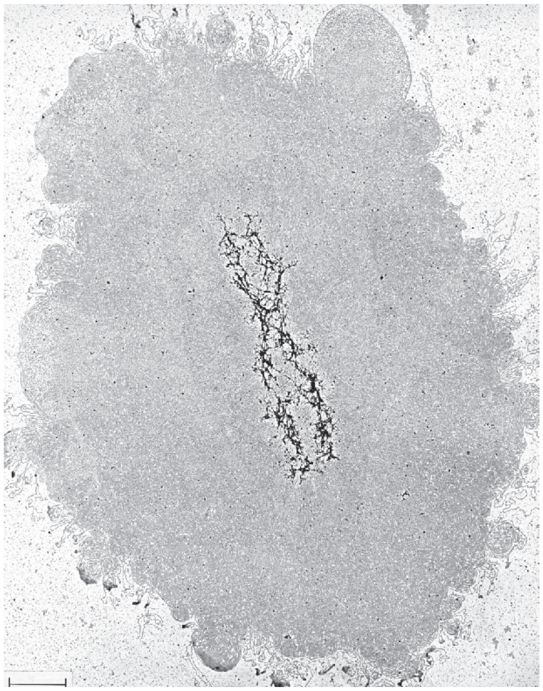


**Figure 15.12** Scaffold protein. When the histones are removed from a eukaryotic chromosome, a fibrous scaffold remains. The DNA loops out from this scaffold. The bar is 2 μm long. (J. Paulson and U. Laemmli, "The structure of histone-depleted metaphase chromosomes," *Cell,* 12:817–28, 1977. Micrograph courtesy of James R. Paulson.)

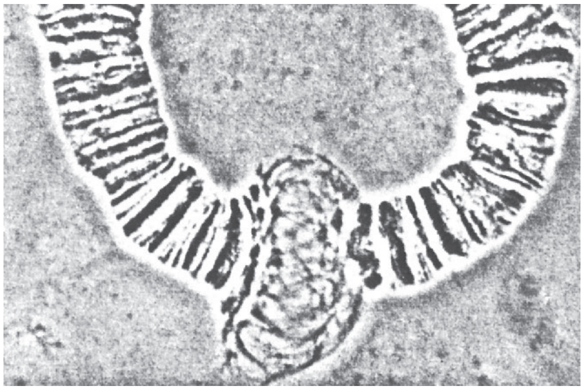


**Figure 15.13** A chromosome puff on the left arm of chromosome 3 of the midge *Chironomus pallidivittatus.* (Jan-Erik Edström, et al., *Developmental Biology* 91:131–37, 1982, Figure 1B, Academic Press.)

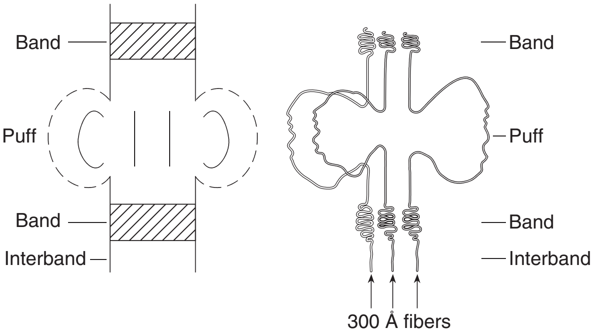


**Figure 15.14** Polytene chromosome with bands and a puff. Three of the approximately one thousand synapsed chromatids are shown diagrammatically on the *right.*

### *Polyteny, Puffs, and Balbiani Rings*

*Drosophila's* salivary glands, as well as some other tissues of *Drosophila* and other diptera, contain giant banded chromosomes (see fig. 6.12) that result from the replication of the chromosomes and the synapsis of homologues without cell division (endomitosis). These chromosomes consist of more than one thousand copies of the same chromatid and appear as alternating dark bands and lighter interband regions. The dark bands are referred to as *chromomeres.* Also seen are diffuse areas called **chromosome puffs** (fig. 15.13). Chromosome puffs are also referred to as **Balbiani rings.** These rings were originally defined as puffs in the midge, *Chironomus,* whose polytene chromosomes were discovered by E. G. Balbiani in 1881. Currently, the term applies to all puffs, or at least the larger puffs, in all species with polytene chromosomes.

The structure of the polytene chromosome can be explained by the diagram in figure 15.14. Dark bands (chromomeres) are due to tight coiling of the 300 Å fiber; light interband regions are due to looser coiling. The figure also shows how chromosome puffs would come about as fibers unfold in regions of active transcription.

Staining with reagents specific for RNA, such as toluidine blue, or autoradiography with tritiated ($^3$H) uridine, have been used to demonstrate that active transcription is going on in the puffs but not in neighboring regions of the polytene chromosomes. The messenger RNA isolated from cells with puffs has also been shown to hybridize only to the puffed regions of the chromosomes. Thus, these regions of the DNA are complementary to the messenger RNA (fig. 15.15) and represent areas of active transcription. Modern recombinant DNA techniques have also shown that many puffs probably represent the transcription of only one gene, although there are exceptions.

Puffs generally fall into four categories. *Stage-specific puffs* appear during a certain stage of development, such as molting. *Tissue-specific puffs* are active in one tissue but not another. (In dipteran larvae, tissues other than the salivary glands, such as the midgut and Malpighian tubules, have polytene chromosomes.) *Constitutive puffs* are active almost all the time in a specific tissue. And *environmentally induced puffs* appear after some environmental change, such as heat shock (fig. 15.16). In *Drosophila,* about 80% of the puffs are stage specific; in *Chironomus,* only about 20% are. For example, at the time of molt in insects, the hormone ecdysone is secreted by the prothoracic gland. At the same time, many puff patterns change (fig. 15.17). Similar changes in puff patterns can be induced by the injection of ecdysone. Hence, molting, a stage-specific developmental sequence, is related to a sequential transcription sequence in the chromosomes.

### Lampbrush Chromosomes

**Lampbrush chromosomes,** which occur in amphibian oocytes, are so named because their looped-out configuration has the appearance of a brush for cleaning lamps, now a relatively uncommon household item (fig. 15.18). The loops of the lampbrush chromosomes are covered by an RNA matrix and are the sites of active transcription. Presumably, the loops are unwindings of the single chromosome, similar to the unwindings in the polytene chromosome shown in figure 15.14. Thus, under certain circumstances, such as in polytene chromosomal puffs

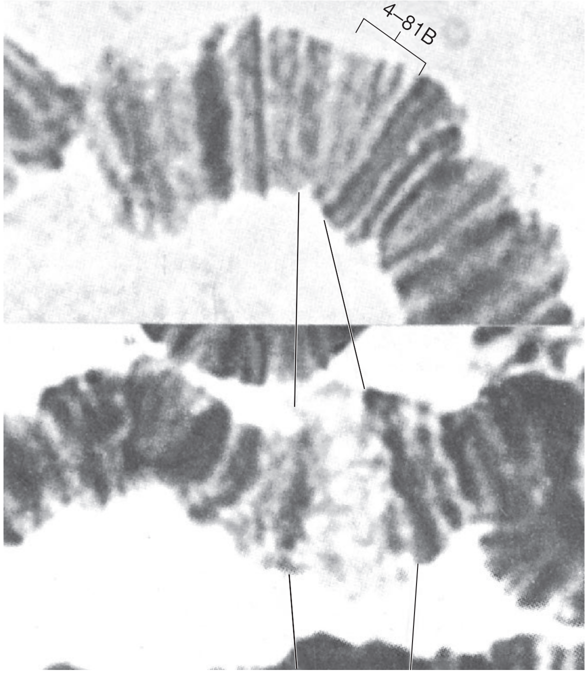


**Figure 15.16** Puff 4-81B of the salivary gland in *Drosophila hydei* is induced by heat shock (37° C for one-half hour). At the *top,* normal activity. At the *bottom,* temperature shock in vitro, resulting in the puff. (Source: H. D. Berendes, et al., "Experimental puffs in salivary gland chromosomes of *Drosophila hydei,*" *Chromosoma* [Berl.] 16:35–46, Fig. 4a–b, 1965. © Springer-Verlag.)

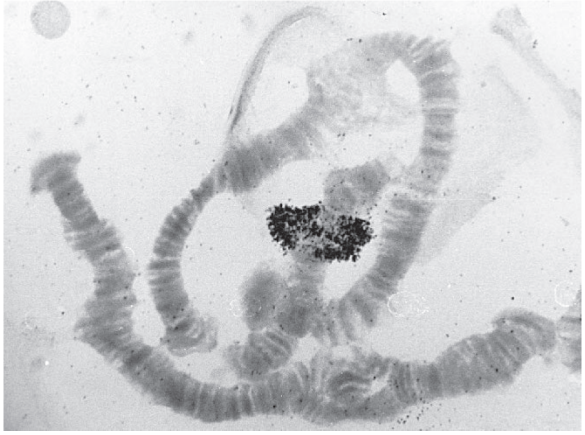


**Figure 15.15** Hybridization at a *Chironomus tentans* salivary gland chromosome puff. The chromosomal DNA is hybridized with labeled RNA (*black dots*) transcribed from the locus. The activity of the locus is forming the puff. (Reprinted by permission from B. Lambert, "Repeated DNA sequences in a Balbiani ring," *Journal of Molecular Biology,* 72:65–75, 1972. Copyright by Academic Press, Inc. (London) Ltd.)

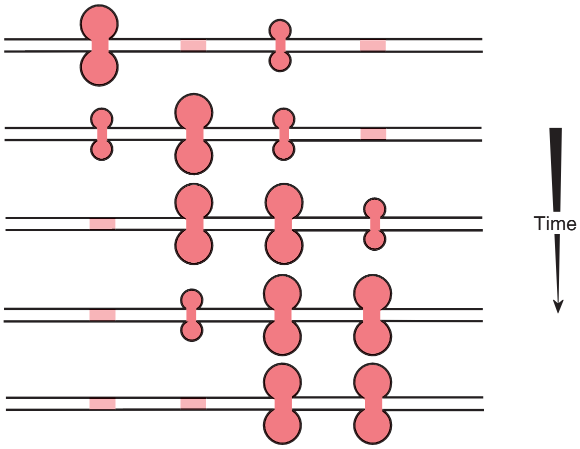


**Figure 15.17** Puff patterns on a segment of a *Chironomus tentans* salivary gland chromosome during molt. As time proceeds, puffs appear and disappear and change in size.

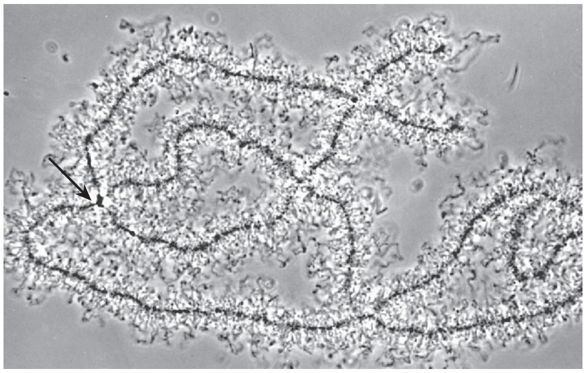


**Figure 15.18** Lampbrush chromosome of the newt, *Notophthalmus viridescens.* Centromere is at the the *left* (*arrow*); the two long homologues are held together by three chiasmata. Magnification 238×. (Source: Joseph G. Gall, figure 2 in D. M. Prescott, ed., *Methods in Cell Physiology,* vol. 2 [New York: Academic Press, 1966], 39. Reproduced by permission.)

and in lampbrush chromosomes, active transcription can be seen in the light microscope. Since only certain bands puff at any one moment in polytene chromosomes, and since the loops of lampbrush chromosomes are of various sizes (with some regions not looped at all), we have evidence of specific transcription. However, we have no indication, so far, of the nature of the control of that transcription.

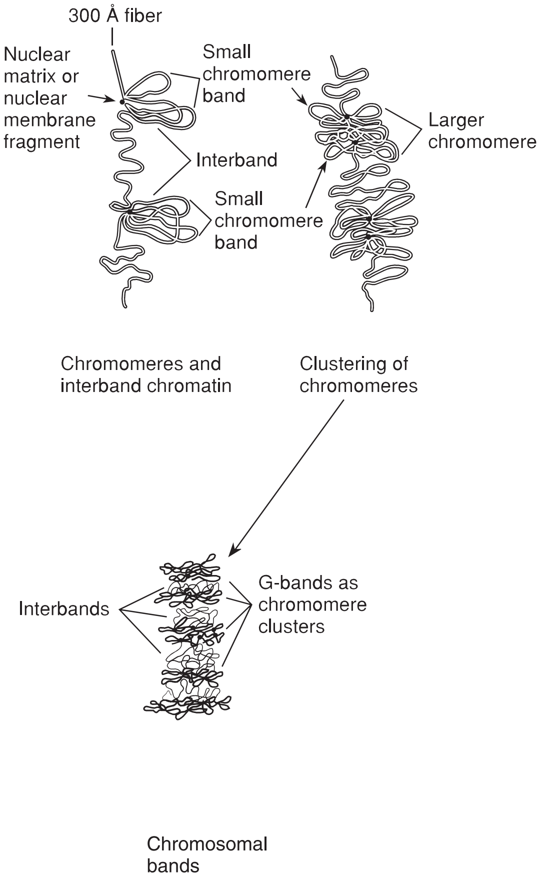

300 Å fiber
Nuclear matrix or nuclear membrane fragment
Small chromomere band
Larger chromomere
Interband
Small chromomere band
Chromomeres and interband chromatin
Clustering of chromomeres
Interbands
G-bands as chromomere clusters
Chromosomal bands

**Figure 15.19** Model of eukaryotic (mammalian) chromosomal banding. G-bands are chromomere clusters, which result from the contraction of smaller chromomeres. These, in turn, result from looping of the 300 Å fiber. (Reproduced with permission, from the *Annual Review of Genetics,* Volume 12, © 1978 by Annual Reviews, Inc.)

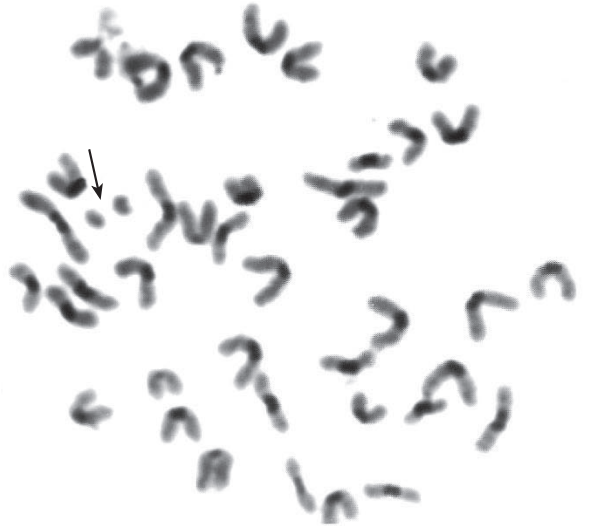
## Chromosomal Banding

Several chromosomal staining techniques reveal consistent banding patterns. By means of these patterns, all of the human chromosomes can be differentiated (see fig. 5.1). Of possibly greater importance is the fact that these staining techniques have provided some insight into the structure of the chromosome. The techniques for staining the C, G, and R chromosomal bands will serve as an illustration.

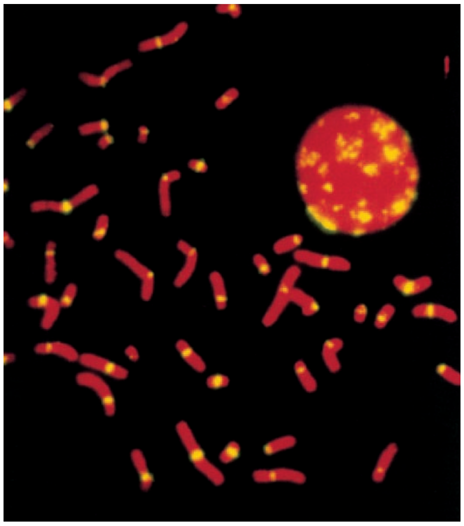
*G-bands* are obtained with **Giemsa stain,** a complex of stains specific for the phosphate groups of DNA. Treatment of fixed chromatin with trypsin or hot salts brings out the G-bands. Giemsa stain enhances banding that is already visible in mitotic chromosomes. The banding pattern is caused by the arrangement of chromomeres. Under careful observation, the major G-bands prove to consist of many smaller chromomeres. This banding appearance has led D. Comings to suggest the mechanism of chromosomal folding shown in figure 15.19.

*C-bands* are Giemsa-stained bands after the chromosomes are treated with NaOH. The *C* is for "centromere," because these bands represent constitutive heterochromatin surrounding the centromeres (fig. 15.20). The DNA is also usually satellite rich. **Satellite DNA** differs in **buoyant density** from the major portion of cellular DNA. When eukaryotic DNA is isolated and centrifuged in CsCl, forming a density gradient, the majority of the DNA forms one band in the gradient at a single buoyant density. The buoyancy is determined by the G-C content of the DNA. However, smaller secondary bands are also usually present, indicating regions of DNA having sequences different from the majority of the cell's DNA (fig. 15.21). DNA isolated this way is referred to as *satellite DNA* because of the secondary, or satellite, bands formed in the density gradient. As we will see, this DNA is found primarily around centromeres and consists of numerous repetitions of a short sequence.

*R-bands* are visible with a technique that stains the regions between G-bands. The chromosomes are fixed, stained with Giemsa, and then viewed with a phase contrast microscope. Since the dark-light pattern is the opposite of the G-band pattern, these bands are called *reverse bands.*



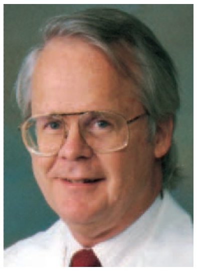
(a)



(b)

**Figure 15.20** (*a*) C banding of chromosomes from a cell in the bone marrow of the house mouse, *Mus musculus.* The *arrow* indicates that the Y chromatids have already separated into two chromosomes. (*b*) Yellow fluorescence indicates a satellite DNA probe in human chromosomes (centromeres). ([*a*] B. Vig, "Sequence of centromere separation: Role of centromeric heterochromatin," *Genetics,* 102:795–806, 1982. [*b*] Photograph Courtesy of Oncor, Inc. Gaithersburg, Maryland.)

From the information gleaned from these staining techniques, D. Comings distinguished between three basic chromatin types: euchromatin, constitutive heterochromatin, and intercalary heterochromatin (table 15.3). Presumably, the only chromatin involved in transcription is **euchromatin. Constitutive heterochromatin** surrounds the centromere and is rich in



David E. Comings (1935– ). (Courtesy Dr. David E. Comings.)

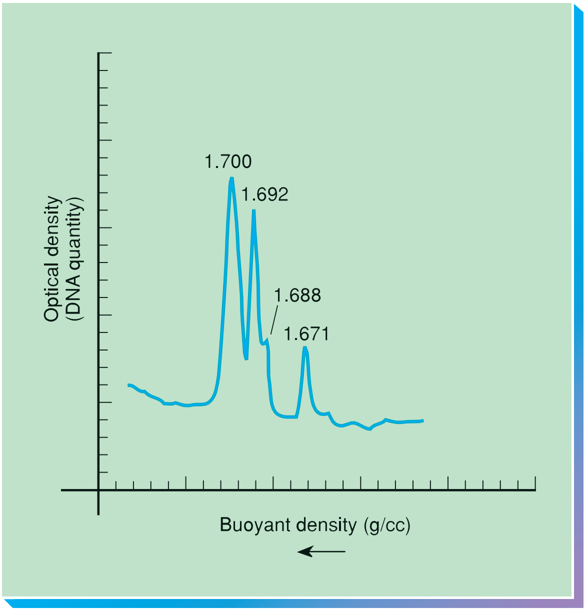


**Figure 15.21** Satellite DNA in *Drosophila virilis.* The quantity of DNA is graphed against the buoyant density (g/cc), resulting in four peaks. The large peak (at *left*) is the major DNA component of the cell; the other three bands are satellite DNA. The left-most of the satellite peaks (1.692) is DNA with a repeating sequence of ACAAACT; the middle satellite peak (1.688) is a sequence of ATAAACT; and the right-most satellite peak (1.671) has a sequence of ACAAATT. (From Joseph G. Gall, et al., *Cold Spring Harbor Laboratory Symposia on Quantitative Biology,* 38:417–21. Copyright © 1974 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. Reprinted by permission.)

**Table 15.3** The Three Major Types of Chromatin in Eukaryotic Chromosomes

| | Euchromatin | Centromeric Constitutive Heterochromatin | Intercalary Heterochromatin |
|---|---|---|---|
| Relation to bands | In R-bands | In C-bands | In G-bands |
| Location | Chromosome arms | Usually centromeric | Chromosome arms |
| Condition during interphase | Usually dispersed | Condensed | Condensed |
| Genetic activity | Usually active | Inactive | Probably inactive |
| Relation to chromomeres | Interchromomeric | Centromeric chromomere | Intercalary chromomeres |

satellite DNA. **Intercalary heterochromatin** is found throughout the chromosome. Thus, it becomes apparent that the eukaryotic chromosome is a relatively complex structure.

## Centromeres and Telomeres

### Centromeres

Two regions of the eukaryotic chromosome have specific functions—the centromere and the telomeres. The centromere is involved in chromosomal movement during mitosis and meiosis, whereas the telomeres terminate the chromosomes. As we pointed out in chapter 3, the terms *centromere* and *kinetochore,* while occasionally used interchangeably, are distinct. The kinetochore is the interface between the visible constriction in the chromosome (the centromere) and the microtubules of the spindle. The kinetochore of higher organisms (e.g., mammals) contains proteins and some RNA. Microscopically, it is a trilaminar structure, attached to chromatin at the inner layer and to microtubules at the outer layer (fig. 15.22).

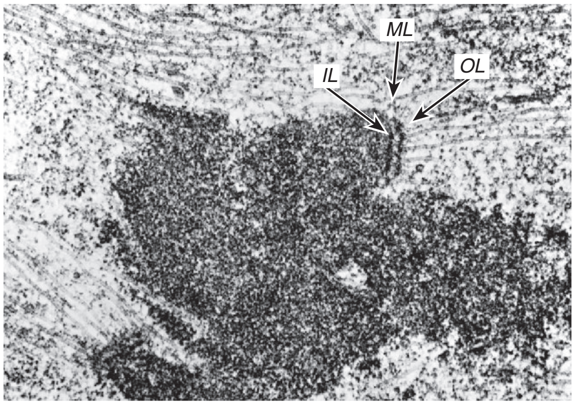


**Figure 15.22** The kinetochore of a metaphase chromosome of the rat kangaroo. *IL, ML,* and *OL* refer to *inner, middle,* and *outer layers,* respectively, of the kinetochore. Note the microtubules attached to the kinetochore and the large mass of dark-staining chromatin making up most of the figure. Magnification 30,800×. (From B. R. Brinkley and J. Cartwright, Jr., *J. Cell Biology,* 50:416–31, 1971.)

Most of our knowledge of the genetics of centromeres has come from work in yeast *(Saccharomyces cerevisiae).* Cells did not maintain most artificially created yeast plasmids because they were lost during mitosis. However, plasmids were isolated that did replicate normally during cell division. Presumably, they contained centromeres, allowing them to replicate and move in synchrony with the host's chromosomes. Further genetic engineering made it possible to isolate smaller and smaller regions that could serve as centromeres. After sequencing the centromeres of fifteen of the sixteen yeast chromosomes, it was possible to conclude that the centromere from yeast is about 250 base pairs long with three consensus regions (fig. 15.23); we are defining a centromere as a sequence of DNA called the *CEN* locus or CEN region. Recent data indicate that this region may contain a single, modified nucleosome associated with region II. The 250 base-pair length of the CEN regions of yeast chromosomes is about 200 Å, the same as the diameter of a microtubule, indicating that only one microtubule attaches to each centromere during mitosis or meiosis in a yeast cell. This region is called a **point centromere** (fig. 15.24).

Higher eukaryotes have larger centromeric regions that attach more microtubules. These regions are referred to as **regional centromeres** (see figs. 15.22 and also 3.12). Regional centromeres range from nineteen to one hundred kilobases (kb; 19,000–100,000 bases) with unique and satellite (repeated sequence) DNA that is heterochromatic and may include expressed genes. We know much less about regional centromeres than we do about point centromeres.

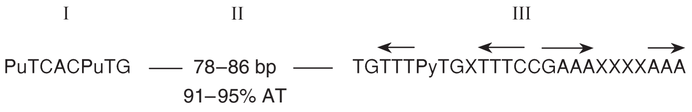


**Figure 15.23** Consensus sequence for the three regions (I–III) of fifteen yeast centromeres. *Pu* represents any purine, *Py* represents any pyrimidine, and *X* represents any base. The *arrows* appear over inverted repeat sequences. (Source: Data from L. Clarke and J. Carbon, "The structure and function of yeast centromeres," *Annual Review of Genetics,* 19:29–56, 1985.)

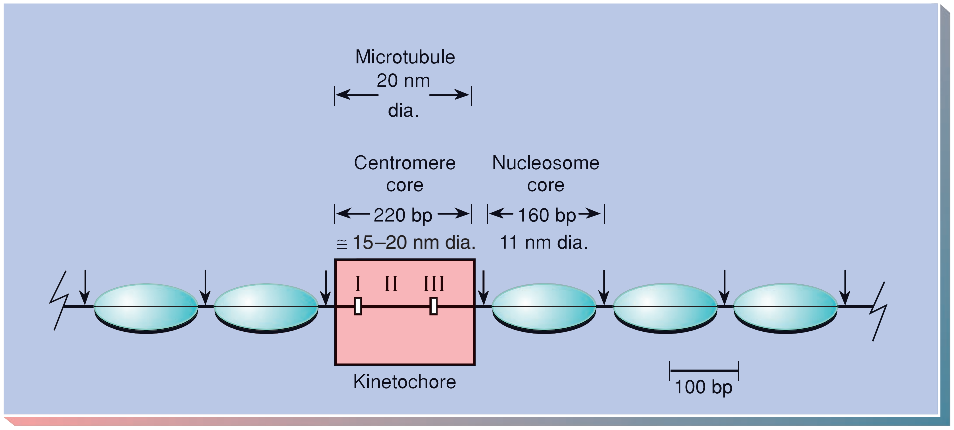


**Figure 15.24** Schematic view of a yeast centromeric region. The *arrows* are the nuclease-hypersensitive sites. A microtubule is about the same width as the centromeric region. (With permission, from the *Annual Review of Genetics,* Volume 19 © 1985 by Annual Reviews www.AnnualReviews.org)

### *Telomeres*

Since eukaryotic chromosomes are linear, each has two ends, referred to as **telomeres,** that not only mark the termination of the linear chromosome but also have several specific functions (fig. 15.25). Telomeres must prevent the chromosomal ends from acting in a "sticky" fashion, the way that broken chromosomal ends act (see chapter 8). In other words, chromosomal ends must not elicit a DNA repair response (see chapter 12). Telomeres must also prevent the ends of chromosomes from being degraded by exonucleases and must allow chromosomal ends to be properly replicated.

Most telomeres isolated so far are repetitions of sequences of five to eight bases. In human beings, the telomeric sequence is TTAGGG, repeated 300 to 5,000 times at the end of each chromosome. The human telomere was discovered by R. Moyzis and his colleagues when they probed the highly repetitive segment of human DNA. (Highly repetitive DNA, as its name implies, consists of numerous copies of a single sequence and usually comprises the satellite components of the cell's DNA; see next section.) When a probe for this sequence was applied to human chromosomes, the sequence was found at the tip of each chromosome in roughly the same quantity (fig. 15.26). This is a highly conserved sequence, found in all vertebrates studied as well as in unicellular trypanosomes. Similar sequences are found in various other eukaryotes (table 15.4); the first sequence was isolated by E. Blackburn and J. Gall in 1978.

When a linear DNA molecule is replicated, the 3′ → 5′ strand can be replicated to the end (see chapter 9). The 5′ → 3′ strand, however, is replicated with RNA primers that are then degraded, leaving a short gap on the progeny strand (fig. 15.27). It is always the G-rich strand of telomeric DNA that ends up single-stranded, forming a 3′ overhang of twelve to sixteen nucleotides. Thus, the normal replication process of a linear DNA molecule leaves an incomplete terminus. Hence, scientists suspected that there would be a unique mechanism for the replication of telomeres.

Telomeric sequences appear to be added de novo without, DNA template assistance by an enzyme called **telomerase,** discovered by E. Blackburn and her colleagues. This was seen when telomeres from another species were engineered into yeast cells. After a cell cycle, the yeast telomeric sequence had been added on at

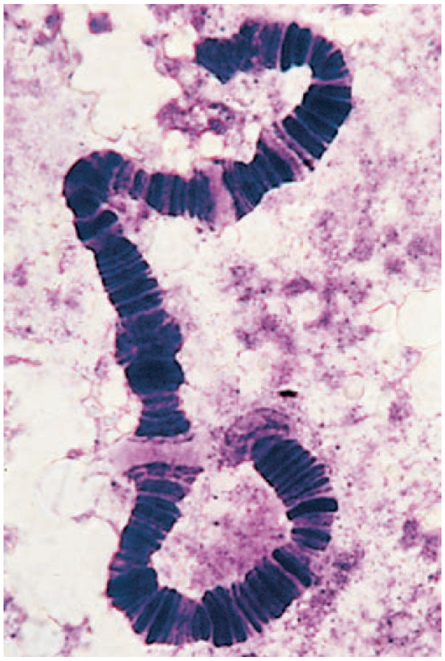


**Figure 15.25** Polytene chromosome from the salivary gland of a *Drosophila* larva showing bands, interbands, puffs, and telomeres. (© David M. Phillips/Visuals Unlimited.)

Elizabeth H. Blackburn (1948– ). (Courtesy of Dr. Elizabeth H. Blackburn.)

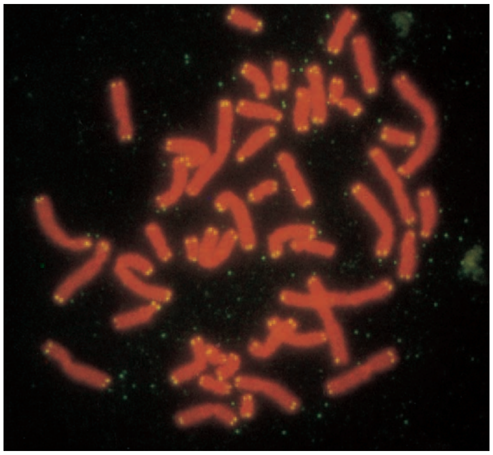


**Figure 15.26** The human genome probed for the telomeric sequence, TTAGGG, using fluorescent staining techniques. The *yellow dots* at the tips of the chromosomes are the probes. (From Robert K. Moyzis, et al., *Proceedings of the National Academy of Science, USA,* 85:6622–26, 1988. Figure 4, left.)

**Table 15.4** **Telomeric Sequences in Eukaryotes; The G-Rich Strand of the Double Helix Is Shown**

| Organism | Telomeric Repeat |
|---|---|
| Human beings, other mammals, birds, reptiles | TTAGGG |
| Trypanosomes | TTAGGG |
| Holotrichous ciliates (*Tetrahymena*) | GGGGTT |
| Hypotrichous ciliates (*Stylonychia*) | GGGGTTTT |
| Yeast | GT, GGT, and GGGT |
| Plants | TTTAGGG |

the ends of the foreign chromosome, the result, presumably, of the telomerase enzyme.

When Blackburn and her colleagues isolated telomerase, they discovered that a segment of RNA, about 160 base pairs, is an integral part of the enzyme. That RNA has a region that is complementary to the G-rich repeat of the telomeric DNA sequence of the species. After careful experimentation, including modifying the gene for the telomerase RNA, Blackburn and her colleagues concluded that telomerase uses its RNA as a template for adding telomeric repeats to the ends of chromosomes. Telomerase is thus a reverse transcriptase, using RNA nucleotides as a template to polymerize DNA nucleotides.

Blackburn and her colleagues proposed that the first step in telomere extension is hybridization of the 3′ end of the telomere with the RNA component of telomerase (fig. 15.28*a*). Then, with the telomerase RNA as a template, the 3′ end of the telomere is extended (fig. 15.28*b*). Finally, a translocation step takes place that displaces the telomere in respect to the RNA, returning to the configuration at the beginning of the process (fig. 15.28*c*). The single-stranded C-rich strand is then synthesized with DNA polymerase and DNA ligase.

Once telomeres have been added to the ends of eukaryotic chromosomes, different organisms use any of three different methods known to protect the ends of

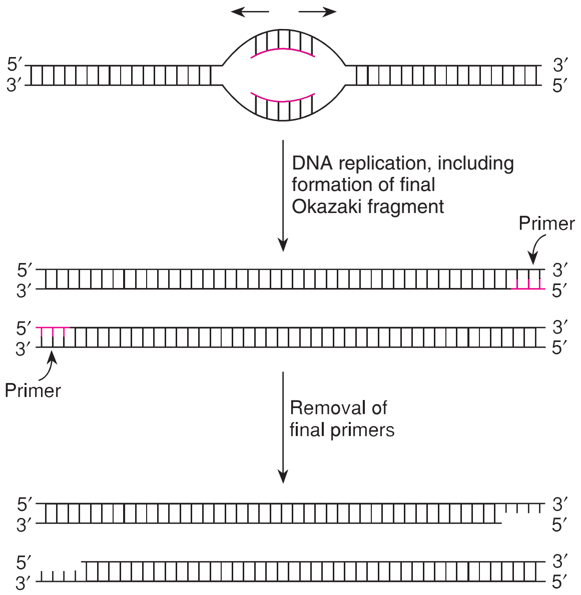

5′
3′
3′
5′
DNA replication, including formation of final Okazaki fragment
Primer
Primer
Removal of final primers

**Figure 15.27** Removal of final primers after the replication of linear DNA creates single-stranded ends.

the chromosomes. First, the guanine-rich DNA can form complex structures. Biochemists have discovered that four guanines can form a planar **G-tetraplex,** with the four bases hydrogen bonded to each other (fig. 15.29). Several structures have been hypothesized to explain the novel ends of these chromosomes (fig. 15.30). Second, proteins have been discovered that bind to the 3′ ends of telomeres. In the ciliate *Oxytricha nova,* a protein called the telomere end-binding protein (TEBP) attaches to the 3′ ends of telomeres and protects them (fig. 15.31). Finally, a novel structure called the ***t-loop*** has been discovered at the end of mammalian telomeres. This loop forms at the ends of chromosomes under the direction of a protein called TRF2 (*t*elomere *r*epeat-binding *f*actor), which causes the 3′ end of the chromosome to loop around and interdigitate into the double helix, forming the loop (fig. 15.32).

How do cells keep track of the number of their telomeric repeats? Proteins have been isolated that bind to telomeres (Rap1 in *Saccharomyces cerevisiae,* TRF1 in human beings). By mutating these proteins or the telomeric sequences, scientists have changed the equilibrium number of telomeric repeats. This led to the current model that the cell counts the number of these proteins bound to the telomeres, not the number of telomeres directly, to know whether telomeres should be added. This is a very active area of research.

In yeast, protozoa, and other single-celled organisms, telomerase is active, keeping the ends of the chromosomes at the appropriate lengths. These cells can divide potentially forever. However, in most cells of higher organisms, telomerase is not active, and the ends of the chromosomes get shorter with each cell division. At a certain telomeric length, the cells no longer divide. However, if telomerase becomes active, and the ends of the chromosomes lengthen, a signal is conveyed to keep cells dividing, which can lead to cancerous growth. In fact, human telomerase was isolated from an immortal cell line (HeLa) derived from cervical cancer cells. Thus,

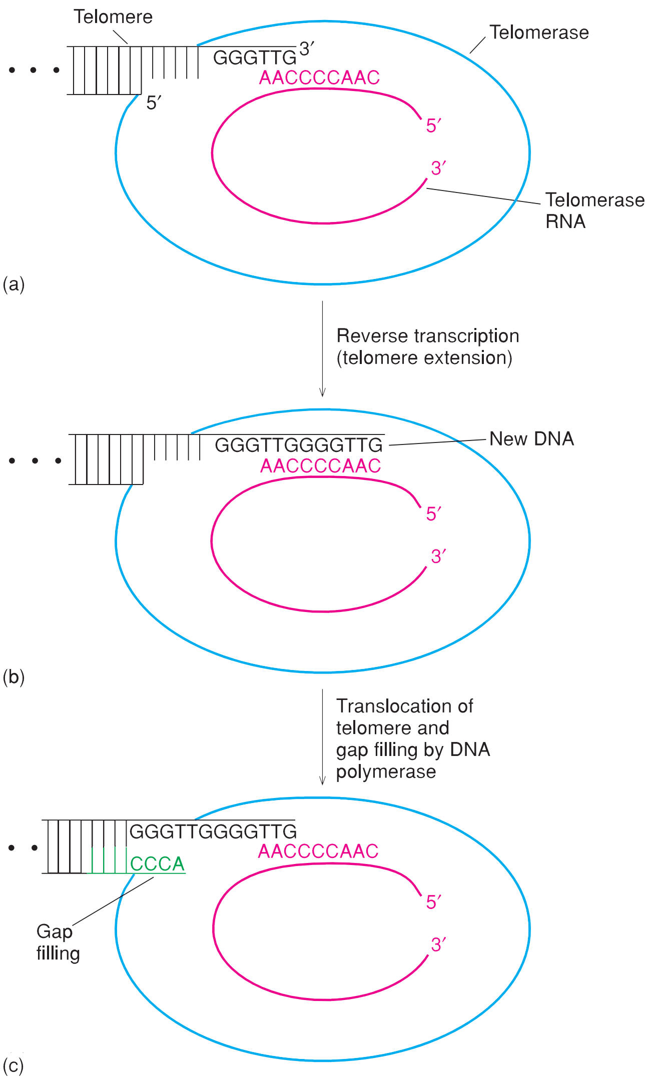

Telomere
Telomerase
GGGTTG 3′
AACCCCAAC
5′
5′
3′
Telomerase RNA
(a)
Reverse transcription (telomere extension)
GGGTTGGGGTTG
AACCCCAAC
New DNA
5′
3′
(b)
Translocation of telomere and gap filling by DNA polymerase
GGGTTGGGGTTG
CCCA
AACCCCAAC
5′
3′
Gap filling
(c)

**Figure 15.28** Telomerase extends telomeres using telomerase RNA (*red*) as a template. Gap filling (*green*) by DNA polymerase and ligase complete the double helix. (Source: Data from Shippen-Lentz and Blackburn, *Science,* 247:550, 1990.)

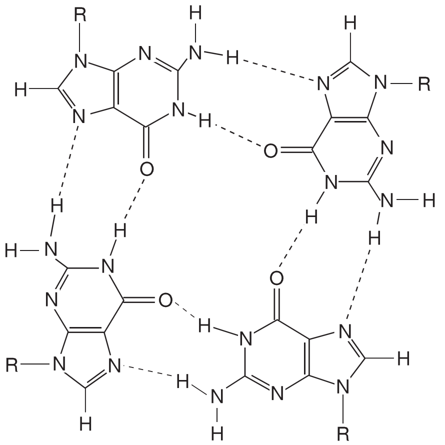


**Figure 15.29** A G-tetraplex can form from four guanines in a plane, hydrogen bonded with each other. (Source: Data from Yong Wang and Dinshaw J. Patel, "Solution structure of the human telomeric repeat d[AG$_3$ (T$_2$AG$_3$)$_3$] G-tetraplex," *Structure,* 1:263–82, December 15, 1993.)

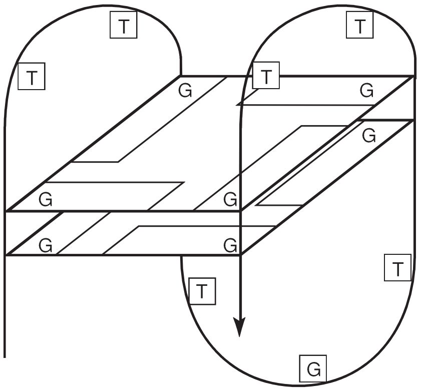


**Figure 15.30** Based on G-tetraplexes (fig. 15.29), the illustrated structure can form at the very tip of a telomere. The sequence d(GGTTGGTGTGGTTGG) is shown forming a four-stranded structure. (Reproduced, with permission, from the *Annual Review of Biophysics and Biomolecular Structure,* Volume 23, © 1994 by Annual Reviews, Inc.)

attention is now turning to the possible clinical application of this knowledge: If telomerase can be deactivated in tumor cells, the cells may stop dividing or die, thereby eliminating the cancer. Further, studying normal telomere shortening, which appears to act as a biological clock, may help us understand the aging process and senescence.

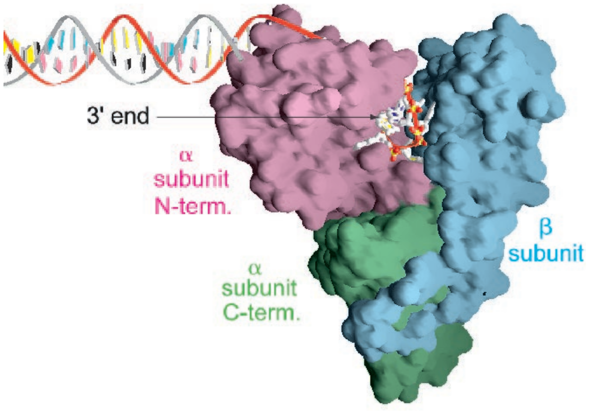


**Figure 15.31** Telomere of the ciliate *Oxytricha nova* shown bound by the dimeric protein called telomere end-binding protein (TEBP). The α and β subunits of the protein form a deep cleft in which the 3′ end of the telomere lies. The folding of the protein into its final form around the DNA may only occur after the DNA has bound, explaining how the DNA could be recognized and placed into such a deep cleft. (Reprinted courtesy of Dr. Martin Horvath.)

## *The C-Value Paradox*

Why do eukaryotes have so much DNA, and why is there huge variation in the DNA content between species of comparable complexity? These questions define the **C-value paradox,** in which *C* refers to the quantity of DNA in a cell. For an example of the paradox, although human beings have 3.3 billion base pairs in the haploid genome, an amoeba has more than 200 billion base pairs. And although an average bony fish has over 300 billion base pairs of DNA in its haploid genome, the Japanese puffer fish has less than half a billion base pairs. If the basic bony fish pattern can be created with less than half a billion base pairs, why does the average bony fish have over 600 times that much DNA? What is this excess DNA doing? To explain the C-value paradox, researchers examined the repetitiveness of DNA, and more recently, probed and sequenced DNA to understand its properties.

### *DNA-DNA Hybridization*

R. Britten and his colleagues, using the technique of **DNA-DNA hybridization,** first systematically analyzed the repetitiveness of the DNA within eukaryotes. When DNA is heated, it denatures or unwinds into single strands; when it cools, it renatures. The rate of renaturation depends on the DNA sequences. If the sample contains DNA with repeated sequences, it will hybridize

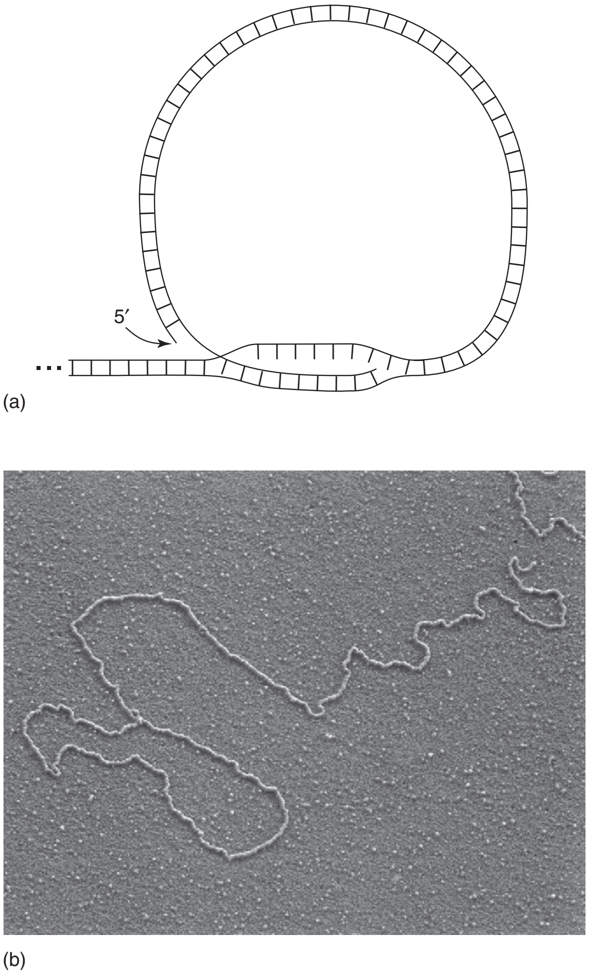
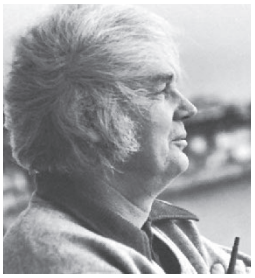


**Figure 15.32** The t-loop at the end of the mammalian telomere. (*a*) A diagram of how the t-loop is formed by the interdigitation of the 3′ end of the telomere into the double helix. (*b*) Electron micrograph of a t-loop from a mouse liver cell. The loop is about 10,000 bases around. ([b] From Jack D. Griffith, et al., "Mammalian telomeres end in a large duplex loop" in *Cell,* 97:503–14, May 14, 1999. Copyright © Cell Press.)



Roy J. Britten (1919– ). (Courtesy of Dr. Roy J. Britten.)

faster than DNA that does not have repeated sequences. From these studies, Britten and his colleagues found that eukaryotic chromosomes contain regions of unique, moderately repetitive, and highly repetitive DNA. **Unique DNA** is, as its name implies, DNA with unrepeated sequences. **Repetitive DNA** is DNA whose sequences are repeated in the genome.

Satellite DNA, found around centromeres (see fig. 15.20), is highly repetitive DNA with a unique repeat length of about two hundred base pairs. Given the quantity of satellite DNA per cell, there must be more than one million repetitions of this two-hundred-nucleotide sequence in higher eukaryotes. At the other end of the spectrum is unique DNA, which makes up most of the transcribed genes of an organism. The rest of the DNA is repetitive DNA in a few to several hundred thousand copies. This repetitive DNA comprises at least three categories. One is "junk" DNA, DNA that is not useful to the organism, made up of untranscribed and parasitic sequences (selfish DNA). Another category is transcribed genes in many copies that have diverged from each other, such as antibody, collagen, and globin genes. We use the term **gene family** to refer to genes that have arisen by duplication, with or without divergence, from an ancestral gene. And finally, transcribed genes in many copies that are virtually identical, such as ribosomal RNA and histone genes, make up a third category of repetitive DNA.

### *Junk DNA*

We saw in chapter 13 that transposons in prokaryotes are generally viewed as selfish or parasitic: They serve no purpose to the cell. The transposons replicate on their own, increasing in number. Eukaryotic transposons are mostly **retrotransposons,** transposable elements that move by way of an RNA intermediate. That is, the retrotransposon is transcribed into RNA and then, by reverse transcription, converted to a cDNA that is then inserted into the genome. These elements can make up 50% of the eukaryotic genome, existing in hundreds of thousands of copies. They generally fall into two categories: LINES and SINES. **Long interspersed elements (LINES),** are up to seven thousand base pairs each and contain genes for reverse transcription, RNA binding, and endonuclease activity. They thus have the ability to jump by way of an RNA intermediate. Human DNA is believed to be composed of about 15% LINES.

**Short interspersed elements (SINES)** are generally derivatives of transfer RNA genes and do not have the ability to retrotranspose on their own. That is, in the past, their transcripts were modified, converted to cDNA by reverse transcription, and then reinserted into the host's genome. They rely on the reverse transcriptase provided by the genes of LINES or retroviruses. One group of SINEs not derived from transfer RNA is derived from the RNA of the signal recognition particle (see chapter 11); members of this group occur in human beings in about five-hundred thousand copies of a three-hundred-base-pair sequence. Because these sequences are cleaved by the restriction endonuclease *Alu*I, they are called the **Alu family.** The human genome is also permeated by remnants of at least a dozen distinct families of ancient retroviruses scattered throughout our chromosomes.

At this point, we can see some potential explanations for the C-value paradox. Much eukaryotic DNA is junk, apparently doing no harm. In some cases, 97% of the host genome is composed of junk DNA. Recent work seems to indicate that gross differences in DNA content between higher organisms may be due to the differing abilities of different species to rid themselves of this parasitic DNA. If it builds up without being removed, the DNA content of the species can soar. Thus, the wide differences in DNA content among higher eukaryotes mentioned at the beginning of this section have little to do with the complexity of the organism, but rather with the ability of the organism to remove junk DNA as it forms.

### Expressed Genes in Many Copies

Several types of genes create a product that is needed in such large quantity that one copy of the gene could not fulfill the cell's needs. We are familiar with the nucleolus, the site of the ribosomal RNA genes (see fig. 10.20). Human beings have about two hundred copies of the major ribosomal RNA gene and about two thousand copies of the 5S ribosomal RNA gene. Fruit flies have about two hundred and one hundred copies, respectively, of the two genes.

In some cases, the normal number of multiple copies of a gene is still not enough. The cell must then resort to **gene amplification,** a process whereby the cell increases the number of copies of the gene. For example, during oogenesis, ribosomal RNA genes (rDNA) are often amplified. In *Xenopus,* rDNA is amplified about one thousand times, which allows an oocyte to accumulate about $10^{12}$ ribosomes. The amplified DNA is in the form of small, circular, extrachromosomal molecules of DNA. Several models have been proposed as to how cells actually amplify their DNA. One model relies on unequal crossing over (as in *Bar* eye in *Drosophila*), whereas another model is based on unscheduled extra DNA replication in a region, followed by recombinational events that generate linear and circular forms of the excess DNA. It is not presently clear which model is correct.

In addition to ribosomal RNA genes, other genes are repeated, ensuring adequate gene products. The number and location of repeated genes are usually discovered by hybridization studies using probes, similar to the way that telomeric DNA was shown to be at the tips of the chromosomes (see fig. 15.26). Repeated genes include the genes for transfer RNAs and histones. The average transfer RNA is repeated about a dozen times in *Drosophila.* Human beings have over thirteen hundred copies of transfer RNA genes in the haploid genome. In many species, the five histone genes form a repeated cluster, although each gene is transcribed independently (fig. 15.33), while prokaryotic operons are transcribed as a unit. The arrangement of histone genes may be more complex in higher forms. There are indications that in mammals, histone genes may lie in small groups or even as individual genes.

Several types of genes occur in similar but not identical forms—that is, an original gene was duplicated but, unlike histone or ribosomal RNA genes, the copies diverged in function. These gene families include globin genes, immunoglobulin genes (see chapter 16), chorion protein (insect eggshell) genes, and *Drosophila* heat shock genes.

### The Globin Gene Family

Globins are oxygen-transporting and storage molecules found in animals, some plants, and microorganisms. In higher vertebrates, there are two types of globins: myoglobin, which stores oxygen in muscles, and hemoglobin, found in red blood cells. Myoglobins function as single molecules, whereas hemoglobins occur as tetramers, two each of two protein chains. Evolution in the globin gene family can be traced by comparative studies of globins in different species as well as molecular studies of globins within a species (see chapter 21). Studying hemoglobins has provided a great deal of information on gene expression and evolution. We turn our attention to the globin gene family in human beings.

During human development, four major hemoglobins appear: embryonic hemoglobin, Hb F, Hb A, and Hb $A_2$

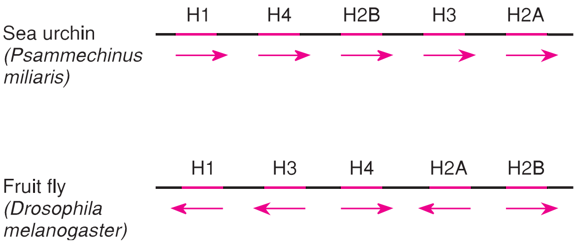


**Figure 15.33** The arrangement of histone genes (*red*) within the five-gene cluster in sea urchins and fruit flies. *Arrows* indicate the direction of transcription. Spacer DNA (*black*) separates the genes.

(table 15.5). Structurally, the ζ (Greek, zeta) subunit (a component of embryonic hemoglobin) is α-like, whereas the rest are β-like (fig. 15.34; see also fig. 10.29). Fetal hemoglobin has a higher affinity for oxygen than does adult hemoglobin, thus allowing fetuses to draw oxygen from their mother's blood. From a comparative study of the DNA sequences, the evolution of the various hemoglobin genes has been inferred (fig. 15.35).

The α genes are located in a cluster on chromosome 16; the β genes are located in a cluster on chromosome 11 (fig. 15.36). These two clusters provide a clear case history of gene duplication, presumably by unequal crossing over, followed by divergence. Having a second or third copy of a gene allows one of the duplicates to diverge (and perhaps to become nonfunctional in the process), whereas the original still performs the required function.

Many diseases of genetic interest involve the hemoglobins. In fact, hemoglobinopathies, including sickle-cell anemia and the thalassemias, are the most common genetic disorders in the world population. The best-known mutation of a hemoglobin gene itself is the one that causes sickle-cell anemia, a mutation of the sixth amino acid of the β chain. In the homozygous state, the disease is usually fatal. However, heterozygotes show an

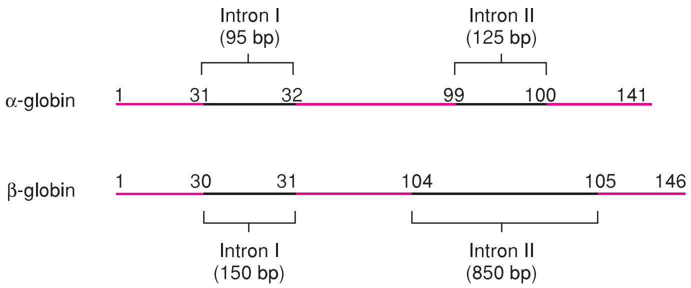

Intron I
(95 bp)
Intron II
(125 bp)
α-globin
1
31
32
99
100
141
β-globin
1
30
31
104
105
146
Intron I
(150 bp)
Intron II
(850 bp)

**Figure 15.34** The structure of adult human α- and β-globin genes. The *numbers* refer to amino acids (or translated codons).

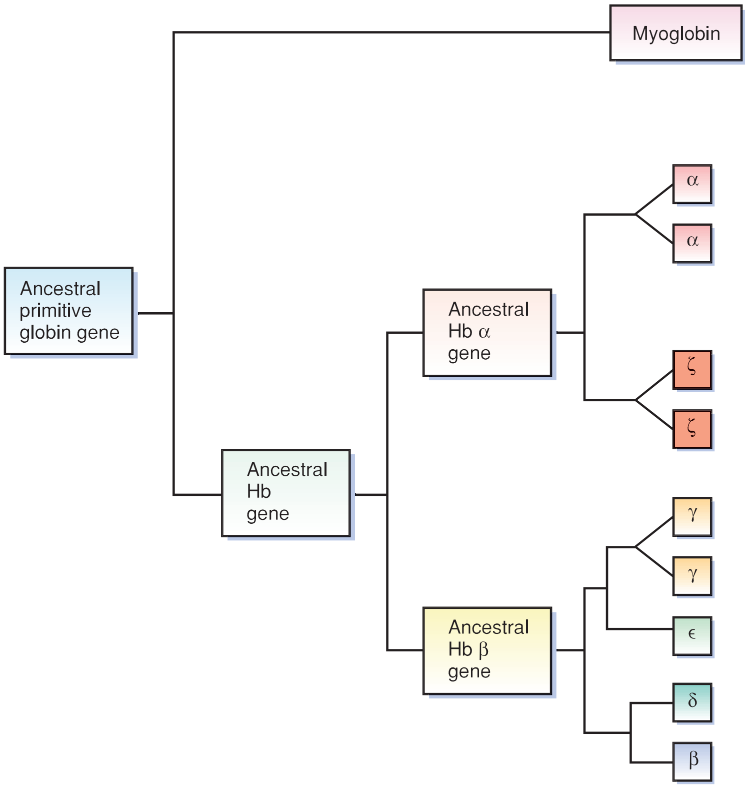

Myoglobin
α
α
Ancestral primitive globin gene
Ancestral Hb α gene
ζ
ζ
Ancestral Hb gene
γ
γ
ϵ
Ancestral Hb β gene
δ
β

**Figure 15.35** The presumed evolution of the various human globin genes from an ancestral primitive gene. The diagram represents a branching tree that begins on the *left* and progresses to the *right.* Each branch point is an evolutionary step in which the genes presumably were duplicated and then either diverged or simply endured as duplicates, as in present-day genes (on the *right*).

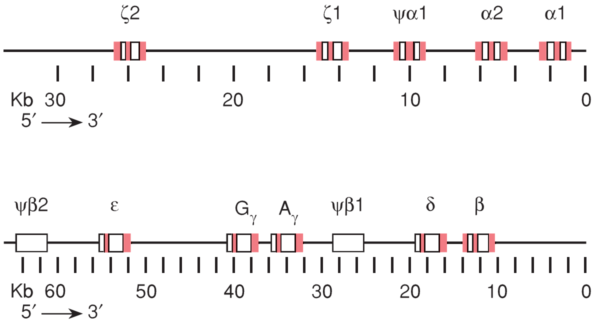
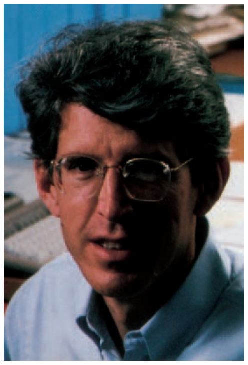


**Figure 15.36** The α- and β-globin gene clusters in human beings. The ψβ1 and 2 and the ψα1 refer to nontranscribing genes (pseudogenes). Mutation has rendered the pseudogenes inactive. Within each gene box, *solid color* refers to exons and *open regions* refer to introns. (Reproduced, with permission, from the *Annual Review of Genetics,* Volume 14, © 1980 by Annual Reviews, Inc.)

**Table 15.5** **Types of Human Hemoglobin**

| Type | Generally When Present | Composition |
|---|---|---|
| Embryonic | Up until eight weeks of gestation and beyond | $\zeta_2\epsilon_2$ |
| Fetal (Hb F) | Eight weeks to birth | $\alpha_2\gamma_2$ |
| Adult (Hb A) | Just before birth and beyond | $\alpha_2\beta_2$ |
| Adult (Hb $A_2$) | In immature cells | $\alpha_2\delta_2$ |

*Note:* Subscripts refer to the numbers of subunits present.



Tom Maniatis (1943– ).
(Courtesy of Dr. Tom Maniatis.)

increased resistance to malaria. One of the ramifications is that the sickle-cell allele is maintained at relatively high frequencies in malarial regions (see chapter 21).

The *thalassemias* are a group of diseases that affect the regulation of the α and β hemoglobin genes. (*Thalassemia* comes from the Greek for "sea blood," because the disease is best known in individuals living around the Mediterranean Sea.) In α and β thalassemias, the α or β subunit, respectively, is present in very low quantities or entirely absent. Many of the genetic defects are deletions, possibly due to unequal crossing over within the globin gene complexes. T. Maniatis showed that β thalassemia is caused by a mutation in the β-globin gene that disrupts RNA splicing. The body compensates by forming $\gamma_4$ or $\beta_4$ hemoglobin in α thalassemias, or $\alpha_2\gamma_2$ or $\alpha_2\delta_2$ in β thalassemias. These are relatively unsuitable or inefficient responses; the diseases range from very mild to very severe and frequently fatal. More information is needed regarding the control of hemoglobin production in the thalassemias.

## SUMMARY

**STUDY OBJECTIVE 1:** To examine the arrangement of DNA and proteins compromising the eukaraoytic chromosome 440–452

To study developmental control in eukaryotes, we must understand the eukaryotic chromosome, which is uninemic: It consists of one DNA double helix per chromosome. Nucleoprotein is composed of DNA, histones, and nonhistone proteins. The nucleosome, a uniform packaging of the DNA, is made of histones. The majority of the nonhistone proteins create the scaffold structure of the chromosome and are not involved in gene regulation. Presumably, very small quantities of the nonhistone proteins take part in the regulation of transcription.

Core DNA, wrapped around nucleosomes, is separated by linker DNA between nucleosomes. There are regions of DNA, vulnerable to nucleases, that do not contain nucleosomes; these are referred to as nuclease-hypersensitive sites. Nucleosomes generally inhibit transcription. The 110 Å nucleosomed DNA forms a 300 Å fiber by coiling into a solenoidlike structure. Coiling of this fiber presumably forms the thick, 2,400 Å fiber seen in metaphase chromosomes.

**STUDY OBJECTIVE 2:** To look at the nature of centromeres and telomeres in eukaroyotic chromosomes 453–457

The centromere and telomeres are specific functional regions of a chromosome. Centromeres isolated from yeast chromosomes have three consensus areas. Telomeres are tandem repeats of a short (five-base-pair to eight-base-pair) segment. Telomeric sequences are added to the ends of chromosomes by the enzyme telomerase, which uses RNA as a template for adding DNA nucleotides. The number of telomeric repeats varies, declining as a cell ages. Telomeric repeat number may control the ability of a cell to replicate and may be implicated in cancerous growth.

Substructuring in the eukaryotic chromosome is demonstrated by G-, C-, and R-banding techniques. C-bands (constitutive heterochromatin) appear to be around the centromeres. These bands consist primarily of satellite DNA, which seems to have a structural role in the chromosome. G-bands (Giemsa bands) presumably represent intercalary heterochromatin and, also presumably, do not have an active transcriptional role. R-bands (reverse bands) appear between the G-bands and represent intercalary euchromatin, the site of transcribed, structural genes.

**STUDY OBJECTIVE 3:** To analyze the nature of the DNA in eukaryotic chromosomes 457–461

Eukaryotes have very large genomes with huge differences in DNA content between organisms similar in complexity, leading to the C-value paradox. Eukaryotic chromosomes contain both unique and repetitive DNA. Highly repetitive DNA is structural (centromeres, telomeres). Junk DNA is mainly short and long interspersed elements. These SINEs and LINEs are often present in hundreds of thousands of copies and can account for 50% of an organism's DNA. They are retrotransposons, transposons that jump by way of an RNA intermediate. Some functional genes also occur in many copies, such as ribosomal RNA genes, histone genes, and globin genes.

## SOLVED PROBLEMS

**PROBLEM 1:** Why is higher-order chromosomal structure expected in eukaryotes but not prokaryotes?

***Answer:*** The simplest explanation is the difference in amount of the genetic material in prokaryotes and eukaryotes. Since the average human chromosome has several centimeters of DNA, that DNA must be contracted to a size in which it can be moved during mitosis and meiosis without tangling and breaking. Nucleosomes provide the first order of coiling, and then several levels of coiling of the nucleosomed DNA bring it down to a manageable size for nuclear divisional processes.

**PROBLEM 2:** Why might we expect to see chromosomal puffs that are tissue- and stage-specific, constitutive, and environmentally induced?

***Answer:*** The various patterns of chromosomal puffing are expected because puffing indicates transcription, the activity of specific genes. Thus, since various tissues are different because they have different proteins, each tissue is expected to have a unique suite of active genes and thus a unique suite of puffs. Similarly, different stages in an insect's development would require different genes to be active, and different puffs should therefore appear at different stages of development. Some genes are active all the time because they specify proteins, such as ribosomal protein genes, that are needed all the time. Finally, environmental insults such as heat shock are known to induce a group of genes that are needed to react to the specific insult, resulting in a suite of puffs that respond consistently to an environmental insult.

## EXERCISES AND PROBLEMS*

### THE EUKARYOTIC CELL

1. Summarize the major differences between eukaryotes and prokaryotes, including the structures of their DNAs.

### THE EUKARYOTIC CHROMOSOME

2. Summarize the evidence that the eukaryotic chromosome is uninemic.
3. What results would you get in the experiment shown in figure 15.1 if the eukaryotic chromosome were not uninemic, but instead had some other number of complete DNA molecules (e.g., binemic)?
4. What are the major protein components of the eukaryotic chromosome? What are their functions?
5. What evidence is used to determine the length of DNA associated with a nucleosome? What is a nuclease-hypersensitive site? What functions are associated with these sites?

---

*Answers to selected exercises and problems are on page A-18.

**6.** What is the protein composition of a nucleosome? What function does histone H1 have?

**7.** What are the relationships among the 110 Å, 300 Å, and 2,400 Å fibers of the eukaryotic chromosome?

**8.** Draw a mitotic chromosome during metaphase. Diagram the various kinds of bands that can be brought out by various staining techniques. What information is known about the DNA content of these bands?

**9.** Give a 300 Å fiber model of the chromosome to account for G-bands.

**10.** Give a 300 Å fiber model of the chromosome to account for polytene chromosomal puffs.

**11.** What are the differences among polytene chromosomes, lampbrush chromosomes, puffs, and Balbiani rings? Draw an example of each.

**12.** Under what circumstances does a chromosomal puff occur? What does it signify?

**13.** What is satellite DNA? What does it signify?

**14.** What is a centromere? a kinetochore? What do we know about the sequences within a yeast centromere?

**15.** What is a telomere? What are its functions? What is its structure?

**16.** Describe three ways in which cells protect their telomeres.

**17.** What functions exist in unique, repetitive, and highly repetitive DNAs?

**18.** How would you use recombinant DNA techniques to locate the number and position of Alu members in the human chromosomes?

**19.** How could you use modern recombinant DNA technology to determine the direction of transcription of the histone genes in figure 15.33?

**20.** How many functional globin genes are there in mammals?

**21.** How could you determine, using modern recombinant DNA techniques, that the α- and β-globin pseudogenes exist?

**22.** Kavenoff and colleagues determined the size of DNA in *Drosophila* chromosomes in two ways: (1) Spectrophotometric measurements were made on the largest intact chromosome. These measurements were then used to calculate the amount of DNA in each chromosome. (2) Nuclei were gently lysed and chromosomes isolated. The lengths of the longest DNA molecules were measured, and those lengths were used to determine the amount of DNA in each molecule. What results for each method would you expect if

**a.** the chromosomes contain one DNA molecule?

**b.** the chromosomes contain more than one DNA molecule?

**23.** What can be said about the base composition of the satellite DNA with a density of 1.671 in figure 15.21?

**24.** When chromatin is partially digested with an endonuclease, the proteins removed, and the DNA separated in a sizing gel, DNA fragments in multiples of two hundred base pairs are found. Provide an explanation for this observation.

**25.** If chromatin is digested with an endonuclease to produce two hundred base-pair fragments, and these fragments are then used for transcription experiments, very little RNA is made. Provide an explanation for this observation.

**26.** Can nucleosomes contain the DNA for one gene? Explain.

**27.** If radioactive probes are made from highly repetitive DNA, these probes hybridize *in situ* mainly to centromeric and telomeric regions. What does this result suggest about the organization of chromosomes?

**28.** Would you expect archaeal species to have nucleosomes?

**29.** What is the C-value paradox, and how is it explained?

**30.** What are the origins of SINEs and LINEs?

## CRITICAL THINKING QUESTIONS

**1.** How could comparative DNA studies aid us in understanding the roles of the different kinds of DNA present in the eukaryotic chromosome?

**2.** How could mutations involving telomeres lead to cancer?

*Suggested Readings for chapter 15 are on page B-14.*

# 16

# GENE EXPRESSION

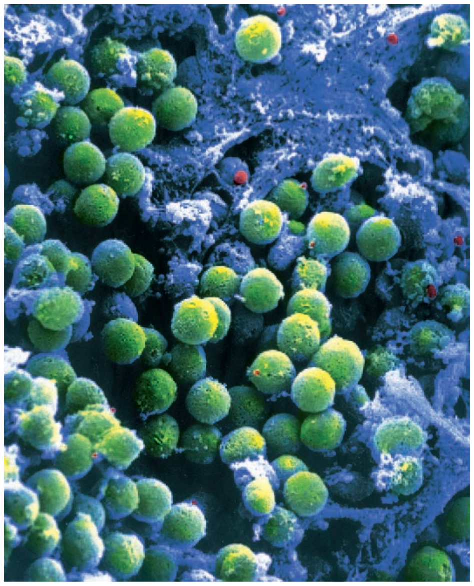
## *Control in Eukaryotes*

## STUDY OBJECTIVES

1. To examine the control of transcription in eukaryotes 465
2. To analyze the genetic control of development in eukaryotes 469
3. To study the mechanisms causing cancer 484
4. To study the genetic mechanisms that generate antibody diversity 492

## STUDY OUTLINE

**Control of Transcription in Eukaryotes** 465
  Chromatin Remodeling 465
  Specific Transcription Factors 466
  Methylation of DNA 466
  Signal Transduction 467
  Transposons 468
**Patterns in Development** 469
  *Drosophila* Development 469
  Developmental Genetics of *Drosophila* 471
  Plants 479
  Other Models of Development 483
**Cancer** 484
  Mutational Nature of Cancer 484
  Viral Nature of Cancer 487
  Environmental Causes of Cancer 492
**Immunogenetics** 492
  Immunoglobulins 493
  Antibody Diversity 494
  T-Cell Receptors and MHC Proteins 498
*Summary* 505
*Solved Problems* 505
*Exercises and Problems* 506
*Critical Thinking Questions* 507
**Box 16.1 Protein Motifs of DNA Recognition** 480
**Box 16.2 Chromosomal Painting** 486
**Box 16.3 AIDS and Retroviruses** 502



Artificially colored scanning electron micrograph of T-lymphocytes, white blood cells involved in the immune system. The cells are seen in the thymus gland, where they mature.

(© CNRI/SPL/Photo Researchers, Inc.)

In this chapter, we turn our attention to the control of gene expression in eukaryotes. We concentrate on the roles that chromatin remodeling, specific transcription factors (transcription activating proteins), and DNA methylation play in determining which genes are expressed at a particular time in a particular cell. We also look at some other possible factors in the control of gene expression: transposons and Z DNA. We then look at the control of gene expression during development, using the fruit fly as a model system. A single cell, the zygote, becomes a whole organism through controlled cascades of gene expression, pathways that are highly conserved in evolution and relatively few in number. Finally, we look at cancer—cell growth out of control—and immunogenetics, the way in which immunological diversity is generated.

## CONTROL OF TRANSCRIPTION IN EUKARYOTES

In prokaryotes, an RNA polymerase holoenzyme with its promoter-recognizing sigma factor is generally active, transcribing at high levels; repressors are needed to prevent transcription. In eukaryotes, an RNA polymerase holoenzyme (e.g., RNA polymerase II), with its promoter-recognizing TFIID, is generally not transcribing; it needs access to the promoter, which is usually wrapped around nucleosomes, and it needs specific transcription factors to become active (fig. 16.1). Thus, although the parts of the transcribing machinery of prokaryotes and eukaryotes are generally similar, the essence of prokaryotic transcription is activity, whereas the essence of eukaryotic transcription is inactivity. In addition, eukaryotes generally do not have operons; however, groups of eukaryotic genes involved in the same pathway or function can be induced simultaneously by having common enhancers that respond to the same specific transcription factors. Such a group of genes is called a **syn-expression group.**

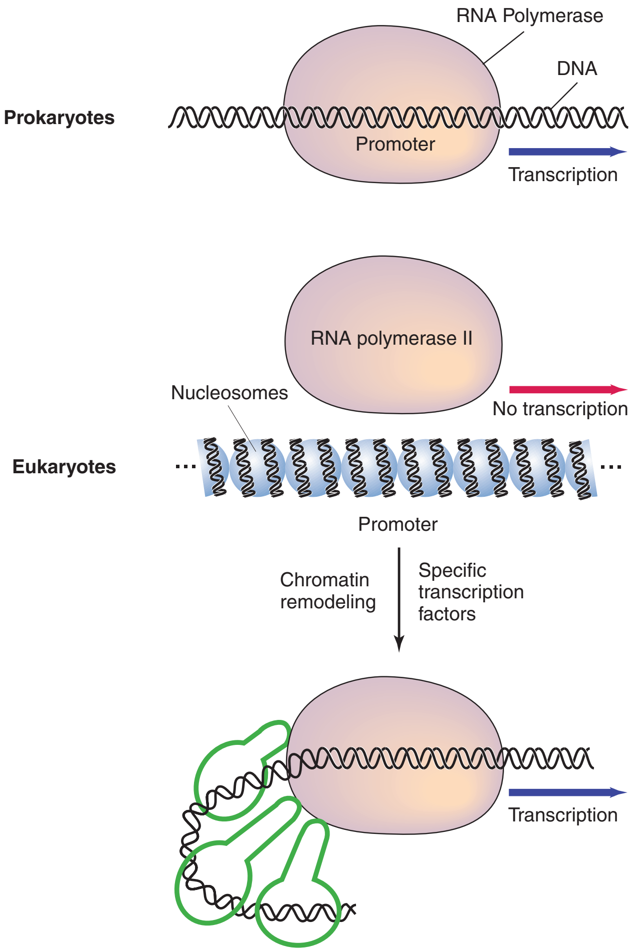


**Figure 16.1** In prokaryotes, the default condition is active transcription. In eukaryotes, the default condition is no transcription since the DNA of promoters is usually wrapped around nucleosomes and specific transcription factors are needed to recruit the polymerase holoenzyme. Transcription in eukaryotes is generally initiated when specific transcription factors bind to enhancer sequences near the promoter, and chromatin is remodeled at the promoter.

### *Chromatin Remodeling*

For transcription to take place in eukaryotes, the DNA must be available for the preinitiation complex to form, with its RNA polymerase and general transcription factors. It appears that DNA wrapped around nucleosomes is often not accessible for the formation of the preinitiation complex, but is available for recognition by transcription-activating proteins, also called specific transcription factors (as compared to the general transcription factors of the RNA polymerase machine; see chapter 10). One model of initiation of transcription by genes whose promoters are wrapped around nucleosomes is for specific transcription factors to recruit chromatin-remodeling proteins. As we discussed in chapter 15, there are two general classes of proteins that remodel nucleosomes: histone acetyl transferases and ATP-dependent chromatin remodeling proteins such as the SWI/SNF complex in yeast. Thus, the presence of one or more specific transcription factors can begin the process of transcription by recruiting chromatin-remodeling proteins that allow the RNA polymerase access to the promoter.

## Specific Transcription Factors

As we discussed in chapter 10, eukaryotic transcription begins with the formation of a preinitiation complex formed by the amalgamation of a group of general transcription factors (such as TFIID in RNA polymerase II formation). Proteins that exert control over transcription at specific promoters are the specific transcription factors (see figure 10.24). These proteins generally have two domains: a domain that recognizes a specific DNA sequence, and a domain that recognizes another protein, such as a protein in the preinitiation complex. Thus, these proteins recognize signals in the vicinity of the promoter of a gene, bind there, and initiate transcription. Currently, we believe that the majority of specific transcription factors act by recruiting the components of the RNA polymerase holoenzyme. Thus, the binding of a specific transcription factor at a promoter is the first step in the formation of a preinitiation complex at the promoter of a gene. Some transcription-activating proteins also recruit chromatin-remodeling proteins.

An example of a specific transcription factor is Dorsal, the product of the *dorsal* gene in fruit flies, active in development. Dorsal controls the transcription of several genes and at several different levels of protein concentration. The ability to have different effects at different concentrations is extremely important, allowing gradients of the same protein to control the expression of different genes. One gene Dorsal controls is *rhomboid,* which has three sites in its promoter that Dorsal binds to, initiating transcription. Another gene, *twist,* also has three sites in its promoter that bind Dorsal, also initiating transcription. However, the *rhomboid* sites are more efficient in binding Dorsal; thus, *rhomboid* is transcribed at lower concentrations of Dorsal than *twist* is (fig. 16.2). One other signal in the control of transcription that is of current interest is methylation.

## Methylation of DNA

The importance of methylation in DNA-protein interactions is well known. In chapter 13, we showed that a particular DNA sequence could be protected from restriction endonucleases if it were methylated. A small percentage of cytosine residues are methylated in many eukaryotic organisms, mainly in CpG sequences (see fig. 13.3); 80% of the cytosines in CpG sequences in human DNA are methylated. (Often, when we refer to a sequence of two bases on the same strand of DNA, we put a "p" between them—CpG—to indicate that they are on the same strand connected by a phosphodiester bond and not on two different strands as a hydrogen-bonded base pair.)

The degree of methylation of DNA is related to the silencing of a gene. Genes that are dormant in one cell type but active in another, or genes that are dormant at one stage of development but active in another, are usually less methylated when active and more fully methylated when inactive. For example, adenovirus, a cancer-causing virus, has been observed in many eukaryotic cell lines. In most lines in which the adenovirus DNA has integrated into the host chromosome, late viral genes are turned off. These genes are highly methylated at their CCGG or GCGC sites.

In addition, chemicals that prevent methylation frequently activate previously dormant genes. For example, 5-azacytidine inhibits methylation; X chromosomal genes, which are normally deactivated, can be reactivated by treatment with 5-azacytidine. There are numerous other examples of the activation of genes after treatment with this chemical. The activated genes lack methylated cytosines that were previously methylated. Finally, the possibility exists that DNA methylation can affect the pattern of chromatin structure.

Recent work has also indicated that the methylation itself may not prevent transcription, but rather may be a signal for transcriptional inactivity. In the thale cress plant, *Arabidopsis thaliana,* a protein named Mom (for *Mo*rpheus

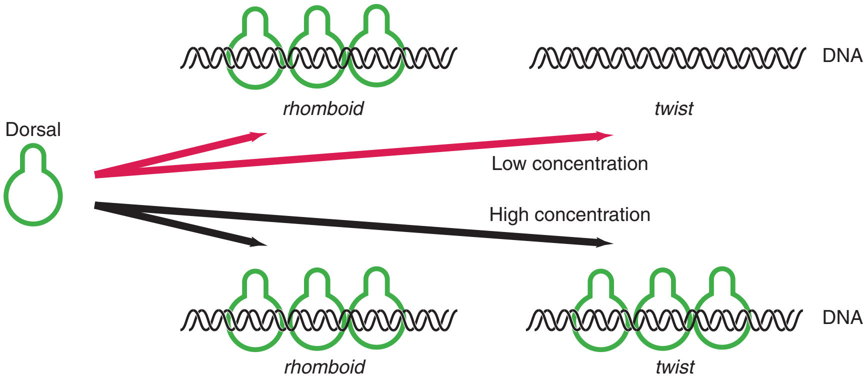


**Figure 16.2** The *rhomboid* and *twist* genes each have three enhancer sequences that are recognized by the Dorsal transcription factor. However, the recognition sequences of the *rhomboid* gene are more efficient at binding Dorsal than the recognition sequences of the *twist* gene. Thus, *rhomboid* is induced at both low and high concentrations of Dorsal, whereas *twist* is induced only at high concentrations.

*m*olecule), has been discovered that, when mutated, results in genes that have heavy methylation levels but are actively transcribed. Thus, the methylation level can be separated from the transcriptional activity of genes, although the two usually occur together. *Arabidopsis* is proving to be a good model in the study of the role of methylation in transcriptional activation because other common model organisms, namely fruit flies, yeast, and the nematode, *Caenorhabditis elegans,* do not have methylation of their DNA.

Further interest has been generated in the role of methylation in controlling gene expression by the discovery of Z DNA, and the fact that Z DNA can be stabilized by methylation (see chapter 9). This observation has led to a model of transcriptional regulation based on alternative DNA structures. Sequences (such as CpG repetitions) that could exist as Z DNA exist as B DNA when being transcribed. If the gene is to be silenced (turned off), the CpG sequences are converted to stable Z DNA by methylation, which then blocks transcription. This possibility has gained some interest because of the recent discovery of an enzyme, double-stranded RNA adenosine deaminase (ADAR1), that binds to Z DNA sequences.

## Signal Transduction

We return to the question of how specific transcription activation factors appear at specific times. As we will describe in the section on development, control of gene expression requires that genes be expressed at specific times and under specific circumstances. If transcription is usually controlled by specific transcription factors, what determines the appearance of these factors at the appropriate times and places? One common mechanism is a **signal transduction pathway,** in which signals pass from the external environment through the cytoplasm, into the nucleus.

For example, in a signal transduction pathway involved in development of the fruit fly, the Toll protein spans the cell membrane (fig. 16.3). It acts as a receptor for the Spätzle protein, which, when detected, causes a change in the cytoplasmic end of Toll, activating it. Activated Toll activates Pelle, a protein kinase that phosphorylates the Cactus protein, causing it to dissociate from Dorsal. Once Dorsal dissociates from Cactus, which acts to repress Dorsal, Dorsal becomes an active specific transcription factor that can cross the nuclear membrane and activate its target gene (fig. 16.3). We thus see that Spätzle attaching to its receptor protein (Toll) on the cell surface results in the activation of the target gene of the Dorsal protein in the nucleus. These pathways can become very complex, with many protein elements. More elements mean more sensitive control of various processes, often requiring that several conditions be met before a gene is activated. In addition, these pathways are usually conserved in evolution. A similar pathway, though more complex, occurs in mammals in which the

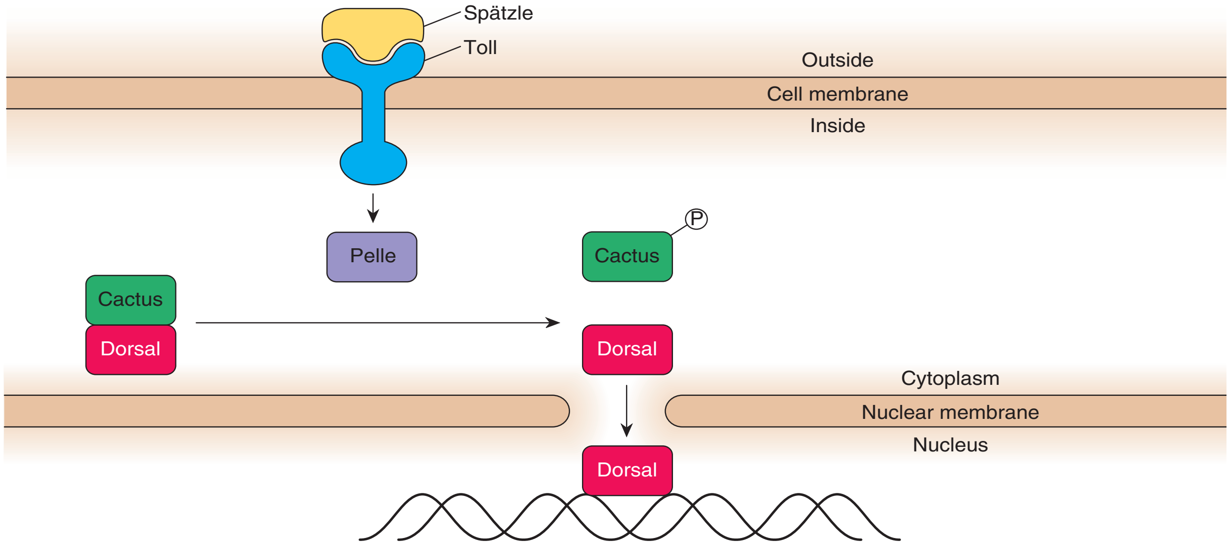


**Figure 16.3** The signal transduction pathway. In this mechanism, the Spätzle protein outside of a cell interacts with the Toll receptor protein, freeing the Dorsal protein to act as a transcription factor in the nucleus. When Toll binds Spätzle, spanning the cell membrane, it changes the configuration of the interior domain of Toll, which then interacts with Pelle, causing it to phosphorylate the Cactus protein. Previously, Cactus had been bound to Dorsal, making Dorsal inactive; phosphorylation of Cactus releases it from Dorsal. Dorsal is then free to cross the nuclear membrane and act as a transcription factor.

target gene is interleukin-1, a protein in the immune system that induces fever. In the mammalian pathway, the signal protein is called Toll-like receptor-4, and the specific transcription factor is called NF-κB.

## Transposons

We have already shown that transposons can affect gene expression in prokaryotes, as, for example, in controlling the flagellar phase in *Salmonella* (see chapter 14). Here we show how transposons can alter or regulate eukaryotic gene expression.

Barbara McClintock discovered transposons in eukaryotes in the 1940s, without the aid of the tools of modern molecular genetics. She won the Nobel Prize for her work in 1983. She observed corn kernels that were streaked or spotted, indicating a high mutation rate. After careful genetic analysis, she showed that the mutability was due to transposons, which she called *controlling elements.*

### *The* Ac-Ds *System*

The *Ac-Ds* system consists of two transposons. McClintock referred to the *Ac (activator)* transposon as an autonomous element and to the *Ds (dissociation)* transposon as a nonautonomous element. *Ds* cannot transpose until *Ac* enters the genome. At that time, *Ds* can transpose, be excised, or cause the chromosome it occurs on to break. *Ds* affects the phenotype by blocking expression of the genes it transposes into, as well as by causing the loss of alleles in acentric chromosomal fragments lost when *Ds* breaks its chromosome.

In figure 16.4, we see three kinds of corn kernels: purple, bronze (light-colored), and bronze with purple spots. The purple kernels result from dominant functioning alleles that provide enzymes in the pathway for purple pigment. In the kernels that are bronze without spots, *Ds* elements have transposed into both copies of the *Bz2* locus, disrupting the pigment pathway. Without the *Ac* element, the *Ds* elements remain in place, and the kernels are a uniform bronze color. In the bronze kernels with purple spots, the *Ac* element has entered the genome in the genetic cross. In the presence of *Ac, Ds* leaves its site in some of the cells, restoring activity to the *Bz2* locus. This restored activity creates purple spots in those cells and in their progeny with the functioning *Bz2* allele (see fig. 14.34*a*). *Ds* and *Ac* elements have been cloned and sequenced. They are typical transposons that are very similar to each other. As might be expected, however, *Ds* has a deletion that prevents it from producing transposase. For *Ds* to transpose, *Ac* must provide the transposase. *Ds* apparently arose from *Ac* by deletion.

It is interesting to note that one of Mendel's original seven characteristics of pea plants, wrinkled peas (*rr:* see fig. 2.3), is caused by a transposon that inserts in the gene for Starch-branching enzyme I. When this gene is functional, the cells produce both branch-chained amylopectins and straight-chained amylose. If the gene fails to produce this enzyme, more sugar is present in the seeds, leading to greater osmotic pressure and, therefore, greater water content. More water is lost from these seeds upon maturation, resulting in greater shrinkage and wrinkling than in the wild-type seeds (*RR* and *Rr*). The transposon that disrupts this gene is about eight hundred base pairs long and is very similar to the *Ds* transposon in maize.

The *Ac-Ds* system disrupts transcription through an invasive element that seems harmful (or at best neutral) to the organism. Mating-type control in yeast, by contrast, is a highly evolved system whose alternative expressions are advantageous to the organism.

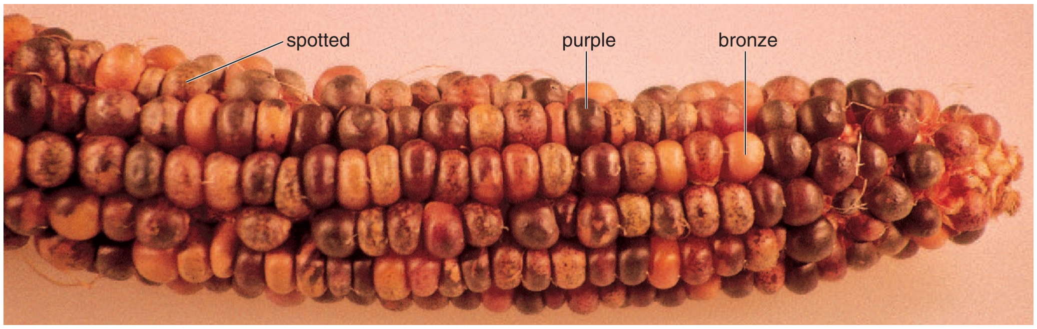


**Figure 16.4** The *Ac-Ds* mutability system in corn. Shown is an ear of corn with purple and bronze kernels. The purple kernels have no transposons. The bronze kernels (light-colored) lack the purple pigment because they have a *Ds* element in both copies of the *Bz2* locus, disrupting pigment production. Without an *Ac* element present, the kernel remains bronze. In the presence of the *Ac* element, the *Ds* element can leave its position, restoring the allele and producing a purple spot in a bronze kernel. Spots differ in size based on when the *Ds* element was excised during the development of the kernel: early excision yields large spots; late yields small spots. (Corn ear courtesy of Dr. Neelima Sinha; Photo by the author.)

### *Control of Mating Type in Yeast*

Transposons determine the mating type in yeast. Haploid yeast cells exist in one of two mating types, **a** and α, determined by the *MAT***a** and *MAT*α alleles. Homothallic strains of yeast switch mating types, as often as every generation. (The term **homothallic,** a misnomer, means that every cell is alike—each can mate with any other. The term was applied before scientists realized that the cells could change mating types.) Homothallism is determined by the dominant *HO* allele that codes for an endonuclease that initiates transposition. Strains that do not change mating type are **heterothallic,** determined by the recessive *ho* allele; no active endonuclease is present to allow transposition, and thus they undergo no change in mating type.

The ability to switch mating types in a single cell implies that both forms of the mating-type gene are present in each cell. In 1971, Y. Oshima and I. Takano proposed that mating type was controlled by a transpositional event, similar to the *Ac-Ds* system in corn or the flagellar phase in *Salmonella.* Later genetic and recombinant DNA studies revealed the exact mechanism.

The third chromosome in yeast contains the mating-type locus (*MAT*). Silent (unexpressed) copies of the mating-type alleles are found on the left and right arms of the same chromosome (fig. 16.5). *HML* contains the silent α allele and *HMR* contains the silent **a** allele. In transposition, a copy of one or the other (*HMR* or *HML*) moves to the *MAT* site, replacing whatever allele was there to begin with. This mechanism has been called a **cassette mechanism.** The *MAT* site is analogous to a cassette player, with *HMR* and *HML* similar to cassette tapes. Transposition brings a new "tape" to the "cassette player."

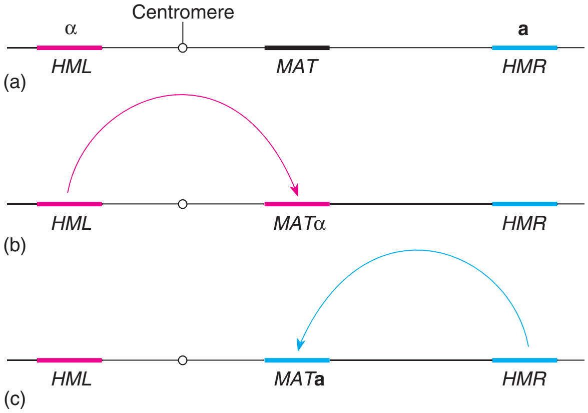


**Figure 16.5** Role of transposition in controlling the mating type in yeast. (*a*) Mating-type loci on the third chromosome. *MAT* is the active mating-type locus. *HML* and *HMR* are silent loci, carrying the two mating-type alleles, α and **a**, respectively. (*b*) Transposition of *HML* to *MAT* results in the *MAT*α allele at the *MAT* site and the α mating type. (*c*) Transposition of *HMR* to *MAT* results in an active *MAT***a** allele, yielding the **a** mating type.

*MAT***a** and *MAT*α each begin a genetic cascade that activates certain genes and represses others. For example, *MAT*α codes for two proteins. The MATα1 protein activates the transcription of an α-factor (a pheromone) gene and an **a**-factor (pheromone) receptor gene. (**Pheromones** are chemical signals, analogous to hormones, that convey information between individuals.) The MATα2 protein represses the **a**-specific genes. Conjugation requires the emission of one type of pheromone and the reception of the other type: An α cell emits α factor and is receptive to **a** factor; an **a** cell emits **a** factor and is receptive to α factor.

In summary, then, transposons can affect eukaryotic gene expression. However, with the exception of a few systems such as mating-type determination in yeast, transposons appear to have a random, disruptive effect on developmental processes.

## PATTERNS IN DEVELOPMENT

**Development** is the orderly sequence of change that produces increasing complexity during the growth of an organism; it is controlled by the differential expression of genes. A central problem of development is explaining **genomic equivalence,** how cells with identical genetic material can give rise to different cell types. A favored approach to understanding the genetic control of development in higher organisms requires first learning the details of the normal developmental process in an organism and then studying the disruption of this normal process by mutation and experimental manipulation.

At one point, scientists believed that development might take place through permanent changes in chromosomes. The idea was that subtle changes might occur in chromosomes during development; these changes would not be observable by karyotyping a cell. Geneticists have explored this hypothesis by several methods. However, the cloning of a mammal, such as the sheep Dolly (see chapter 13), from the cell of an adult demonstrates that adult nuclei are **totipotent:** Any adult nucleus can give rise to the whole organism and all its cellular types, indicating the chromosomes are intact.

### Drosophila *Development*

The fruit fly, *Drosophila melanogaster,* has emerged as an excellent model organism for the study of development. The zygote develops from the egg, in maternal cytoplasm. Maternal messenger RNAs and proteins are the first expressed in the embryo. These substances first determine the broad pattern of the embryo. Then, through signal pathways involving numerous specific transcription factors, they initiate a cascade of gene expression that eventually determines the fate of each cell. As we will see, many parallels exist between the fruit fly and higher organisms.

We will concentrate on two overall patterns of development here: the formation of the basic body plan (anterior-posterior and dorsal-ventral polarity, which results in a segmented embryo that has a front, back, top, and bottom) and the determination of gene expression within segments.

*Drosophila* development begins within a follicle that contains the oocyte surrounded by follicle and nurse cells. The fifteen nurse cells, along with the oocyte, were derived from four divisions of an earlier germ-line cell (fig. 16.6). The nurse cells maintain connections to each other and to the oocyte by cytoplasmic bridges, openings in the membranes surrounding the cells. Thus, the nurse cells can readily pass materials (messenger RNAs and proteins) into the oocyte.

After fertilization, the diploid nuclei divide thirteen times in the space of about 3.5 hours, forming a **syncitium**—a group of nuclei without cell membranes. During this time, most of the nuclei migrate to the inner surface of the developing embryo, where cell membranes eventually form, producing a cellular **blastoderm.** During the syncitial period, materials can move freely through the cytoplasm. At the posterior end of the embryo, several cells, called *pole cells,* that will eventually form the germ cells of the developing fly are set aside (fig. 16.7). Development then proceeds through *gastrulation,* in which cells grow inward, forming the basic germ layers of the embryo (**mesoderm, endoderm,** and **ectoderm**). From these layers, various adult structures will arise. At about six hours of development, furrows become visible in the embryo, delineating segments. The first segments visible are called **parasegments.** They do not give rise to the later segments of the embryo, but rather overlap the later segments in a simple fashion: Each later segment is made up of the anterior end of one parasegment and the posterior end of the next (fig. 16.8). This distinction is meaningful since, as we shall see later, some genes express themselves within the borders of parasegments rather than segments.

The fully segmented embryo has an anterior region, destined to be the head; three thoracic segments, which will give rise to the thorax (the middle region of the fly containing wings and legs); and eight abdominal segments that will give rise to the abdomen. The embryo also has an anterior tip, the **acron,** that will give rise to structures at the very head end—eyes, and antennae; and

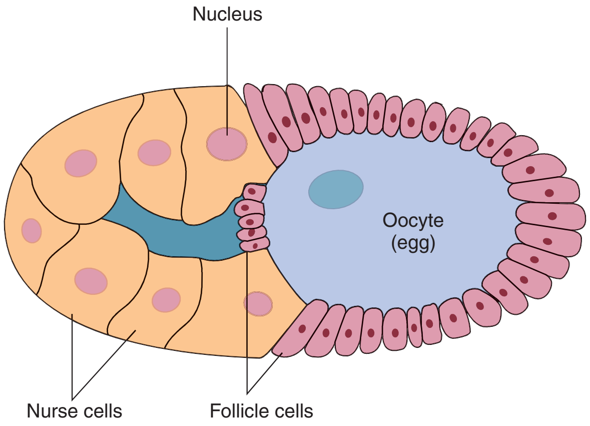


**Figure 16.6** The follicle from a fruit fly, *Drosophila,* consisting of the oocyte, fifteen nurse cells arising from four divisions of a germ-line cell that also gave rise to the oocyte, and follicle cells.

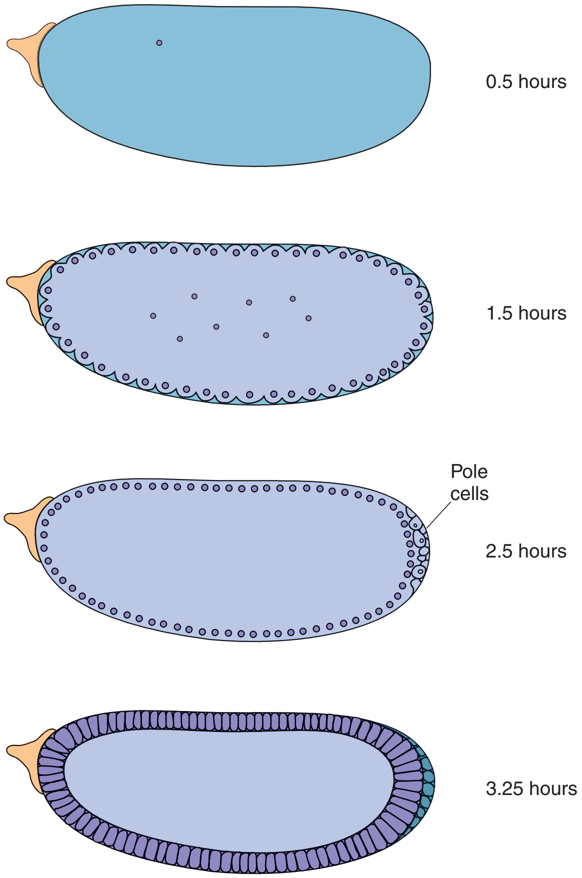


**Figure 16.7** Development of the fertilized *Drosophila* egg after laying. Pole cells, which will be future germ cells, are set apart at about 2 hours. A syncitial blastoderm forms at about 2.5 hours, followed by a cellular blastoderm, consisting of about five thousand cells, at about 3.25 hours.

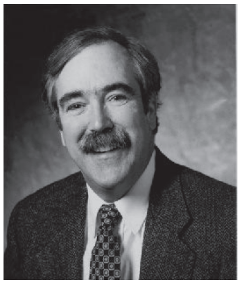
a posterior tip, called the **telson,** that will give rise to the internal structures at the very posterior end of the fly. The fates of these segments have been determined by treating them with various harmless dyes and tracing where the dyes end up. A projection of adult structures on embryonic tissue is called a **fate map.**

## *Developmental Genetics of* Drosophila

### *The General Body Plan*

The role genes play in determining the general axes of the body plan has been worked out at several levels. First, mutations causing female sterility were isolated. (C. Nüsslein-Volhard and E. Wieschaus were instrumental in systematically isolating many of these mutants; they were awarded Nobel prizes for this work.) For example, among normal female flies that were sterile, some

Christiane Nüsslein-Volhard (1942– ). (Courtesy of Christiane Nüsslein-Volhard.)



Eric F. Wieschaus (1947– ). (Courtesy of Dr. Eric F. Wieschaus. Photograph by Denise Applewhite.)

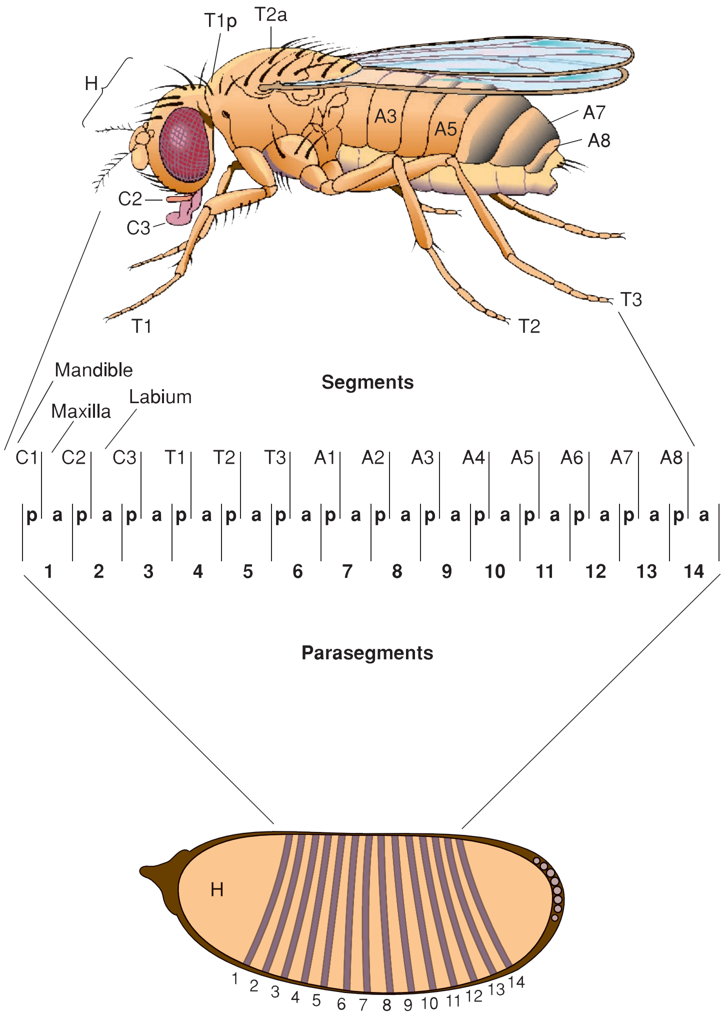


**Figure 16.8** The relationship between parasegments, segments, and the adult fruit fly. The initial segments of the fly are called *parasegments;* the nonsegmented parts of the embryo are called the *acron* at the head end (accounting for eyes and antennae) and the *telson* at the tail end (accounting for the end of the alimentary canal). Later segments are made up of the posterior end of one parasegment and the anterior portion of the next (*p, a*). The later segments map directly on the adult body, accounting for mouthparts (mandible, maxilla, and labium), thoracic segments (1–3), and abdominal segments (1–8). (*H* is for head.) (From P. A. Lawrence, *The Making of a Fly,* Copyright © 1992 Blackwell Science, Ltd., Oxford, England. Reprinted by permission.)

produced embryos without heads or thoracic structures. The gene for this mutation, which has since been cloned and sequenced, is called *bicoid* (fig. 16.9). It codes for a specific transcription factor, the Bicoid protein. (Remember that gene names are italicized, using the first letter, lowercase for recessive and uppercase for dominant; the protein product of these genes is not italicized, but the first letter is capitalized.)

Pricking the anterior end of a normal embryo, causing the loss of cytoplasm from that end (fig. 16.10), can mimic these mutants. This experiment indicates there is some cytoplasmic localization determining the development of the anterior end of the fly. To support that idea further, it was possible to get normal development from a *bicoid* fly by injecting the anterior end with cytoplasm from a normal embryo (fig. 16.10*b*). This process of facilitating normal development by manipulating the embryo is termed a *rescue experiment.* By probing with a complementary oligonucleotide to the *bicoid* messenger RNA, researchers found that the *bicoid* messenger RNA is formed in the nurse cells and then passed into the oocyte, where it becomes localized at the anterior tip (fig. 16.11*a*). After fertilization, this messenger RNA is translated into Bicoid, which begins to diffuse from the anterior end of the egg, until it reaches about 50% of the length of the egg. The protein can be visibly located by treating the eggs with antibodies to the protein; these antibodies can then themselves be made visible (fig. 16.11*b*).

The Bicoid protein is called a **morphogen,** a substance that diffuses through the egg and by its concentration determines the developmental fate of that part of the embryo. Although nurse cells are germ-line cells, they are of maternal origin and not from the embryo. Since maternal cells, not the embryo itself, produce this morphogen, the gene responsible for its production is called a **maternal-effect gene.**

Other maternal-effect genes are involved in formation of the anterior pattern that produces headless embryos. However, they don't appear to produce a morphogen. Rather, these genes seem to be involved in the transport, stabilization, and modification of the morphogen. In mutants of these other genes (*swallow, exuperantia*), Bicoid is found in the nurse cells but not in the embryo; cytoplasm from the nurse cells of these mutants can rescue *bicoid* mutants, indicating that the morphogen is present but not delivered to the oocyte. Only mutants of the *bicoid* gene itself cannot rescue the various headless mutants because only in *bicoid* mutants is the morphogen itself missing.

Through experiments similar to the ones described for *bicoid,* four independent signaling pathways of maternal-effect genes have been isolated. These pathways determine the general body plan of the developing embryo: anterior, posterior, terminal, and dorso-ventral. The posterior pattern is controlled by the gradient of a protein, Nanos. Before the *nanos* gene is active, producing messenger RNA, the first posterior gene active is *oskar;* the localization of *oskar* messenger RNA then defines the localization of *nanos* messenger RNA. Mutant embryos can be rescued by wild-type cytoplasm; the *nanos* mes-

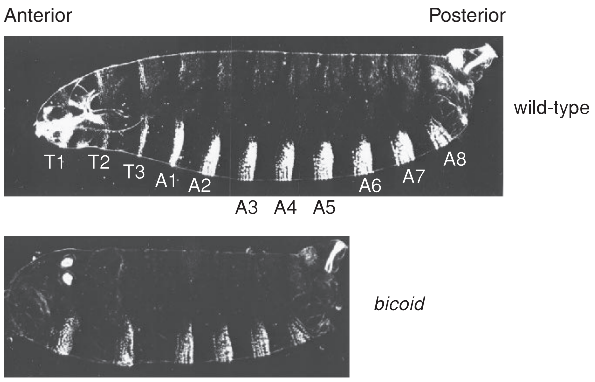


**Figure 16.9** Two *Drosophila* larvae, with cuticular patterns visible on the ventral surfaces. On the *top* is the wild-type with the cuticular pattern coinciding with thoracic and abdominal segments. On the *bottom* is a *bicoid* mutant, lacking head and thoracic structures. (Courtesy of Christiane Nüsslein-Volhard.)

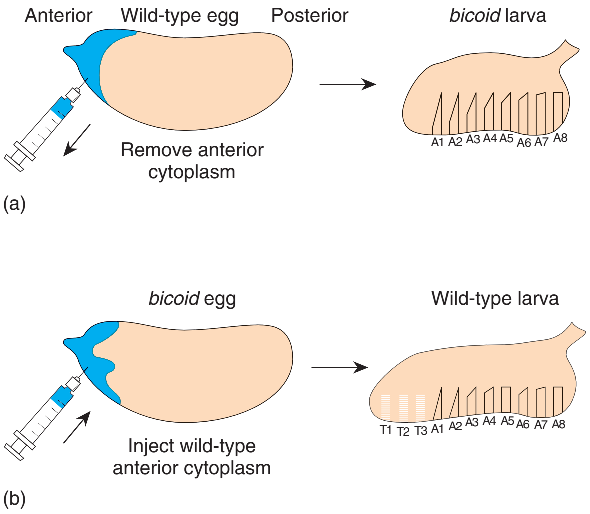


**Figure 16.10** Experiments to demonstrate that a cytoplasmic localization at the anterior end of the fruit fly egg determines anterior structures. (*a*) A wild-type egg has anterior cytoplasm removed, resulting in a larva lacking anterior structures, similar to a *bicoid* mutant. (*b*) A *bicoid* mutant egg has anterior cytoplasm from a wild-type egg injected into the anterior of the egg, resulting in a larva indistinguishable from the wild-type.

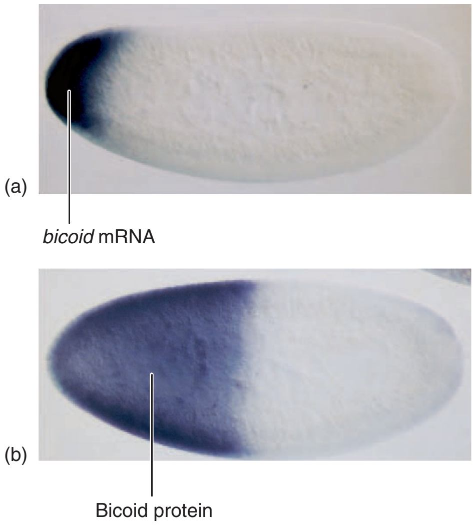


**Figure 16.11** The *bicoid* morphogen first appears in the fruit fly egg as (*a*) messenger RNA at the anterior end of the egg. After fertilization, the messenger is translated into (*b*) Bicoid protein that diffuses toward the posterior end of the embryo. (Courtesy of Daniel St. Johnston.)

senger RNA is localized at the posterior tip of the embryo and produces a protein that diffuses from that tip. Maternal-effect genes that act in a somewhat different manner control the other two pattern systems in the developing embryo.

The terminal pattern controls development of both ends of the embryo; a key gene is *torso.* This gene codes for a membrane-bound tyrosine kinase receptor protein that is found evenly distributed on the outer surface of the developing embryo. (Tyrosine kinases phosphorylate the amino acid tyrosine in specific proteins.) Apparently, other genes in follicle cells located only at the poles of the egg produce a substance that activates the *torso* tyrosine kinase receptor, making it active in only the poles of the egg (fig. 16.12). A maternal-effect gene, *Toll,* that also produces a membrane receptor, controls the dorsoventral axis. Thus, we see that four pathways of maternal-effect genes determine the major body plan of the egg. Two of the pathways are determined by genes that result in diffusion of a morphogen (*bicoid* and *nanos*), and two are determined by genes for membrane receptors (*torso* and *Toll*). About thirty maternal-effect genes are known (table 16.1).

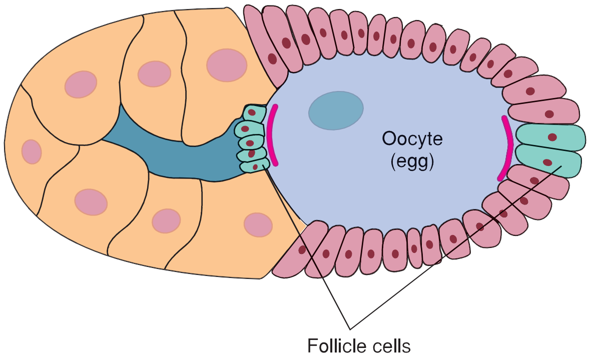


**Figure 16.12** The *Drosophila* follicle, showing follicular cells (*green*) at the tip of the oocyte that secrete a substance that activates the Torso (*torso* gene) tyrosine kinase at the areas marked by *red lines;* the inactivated kinase is located around the surface of the oocyte.

**Table 16.1** **Maternal-Effect Genes in *Drosophila* (Allelic Designations in Parentheses)**

| Anterior | Posterior | Terminal | Dorso-Ventral |
|---|---|---|---|
| *bicoid (bcd)* | *nanos (nos)* | *torso (tor)* | *Toll (Tl)* |
| *swallow (swa)* | *oskar (osk)* | *trunk (trk)* | *nudel (ndl)* |
| *exuperantia (exu)* | *vasa (vas)* | *torsolike (tsl)* | *pipe (pip)* |
| *bicaudal (bic)* | *tudor (tud)* | *polehole [fs(1) ph]* | *windbeutel (wbl)* |
| *Bicaudal-D (BicD)* | *stauffen (stau)* | *Nasrat [fs(1) N]* | *snake (snk)* |
| *Bicaudal-C (BicC)* | *valois (val)* | | *easter (ea)* |
| | *pumilio (pum)* | | *cactus (cact)* |
| | | | *spätzle (spz)* |
| | | | *tube (tub)* |
| | | | *pelle (pll)* |

*Source:* Reprinted with permission from C. Nusslein-Volhard, et al., *Science,* 238:1675–81, 1987. Copyright © 1987 American Association for the Advancement of Science.

Activity of maternal-effect genes in the follicle cells is controlled by an interaction between the oocyte itself and the follicle cells. Follicle cells at the anterior of the oocyte produce *bicoid* messenger RNA as a default condition. At the posterior of the oocyte, the follicle cells produce *nanos* messenger RNA, along with several other gene products. These follicle cells are induced to action by the product of the *gurken* gene in the oocyte; the oocyte nucleus is located posteriorly at this point, and its gene products can be directed to the posterior of the oocyte, where they diffuse to adjacent follicle cells. These cells have a receptor on their surfaces, the product of the *torpedo* gene, that recognizes the *gurken* gene product. Through signal transduction, these follicle cells are induced to express the *nanos* gene (fig. 16.13).

At this point, some product of these follicle cells induces a reorganization of the microtubules in the oocyte, causing the oocyte nucleus to move anteriorly and dorsally. Now, the same *gurken-torpedo* interaction takes place, causing these follicle cells to induce the dorso-

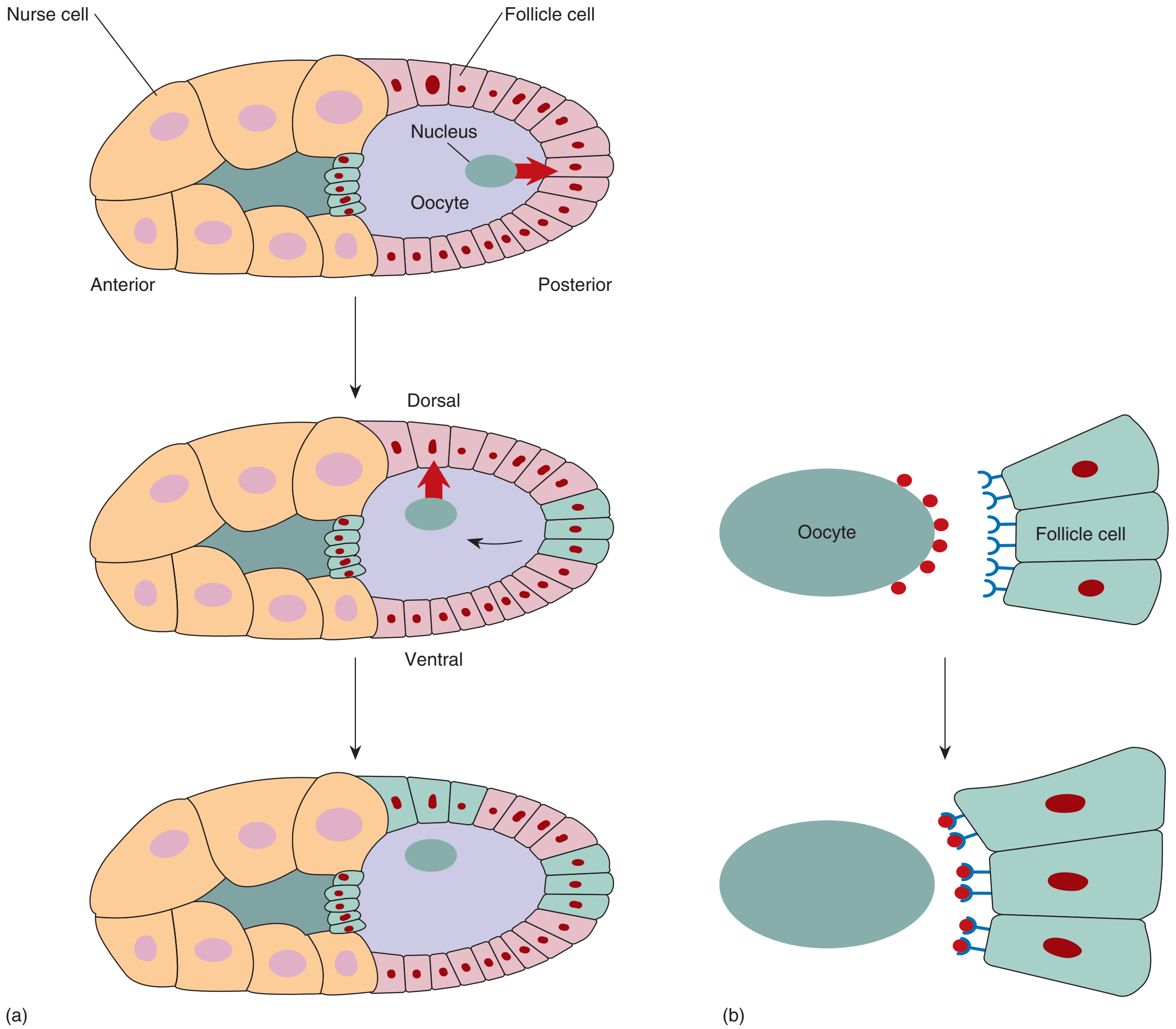


**Figure 16.13** The interaction of the oocyte nucleus and follicle cells early in development. (*a*) The oocyte nucleus, located posteriorly in the oocyte, activates posterior follicle cells. These cells will later provide the *nanos* messenger RNA to control posterior development of the embryo. After this interaction, a product of the follicle cells causes a rearrangement of microtubules in the oocyte, moving the oocyte nucleus anteriorly and dorsally. There the same interaction takes place, in this case activating follicle cells to control dorsal development. (*b*) The oocyte signal (*red circles*) is the product of the *gurken* gene; it interacts with a receptor (*blue Y-shapes*) on the surface of follicle cells, the product of the *torpedo* gene.

ventral axis. As of yet, we don't know all of the signaling going on, nor why two similar cell types react differently to the same oocyte signal (Torpedo), but we do know that maternal-effect genes in the follicle cells are induced by the oocyte itself.

Maternal-effect genes are the first in a series controlling a cascade of gene expression that eventually determines the fates of individual cells in the developing fly embryo. The rest of the genes are zygotic genes, genes active in the cells of the embryo itself. As we move down this cascade of genes, we go from broad patterns to more and more focused gene activity. We go from a single cell with gradients of morphogens to stripes of cells with different genes active.

### *Segmentation Genes*

Once the general body plan of the fly is in place, development continues in the formation of parasegments and then segments. The various organs of the fly's body are produced from these segments. Further development is now under the control of the zygote's own genes, generally referred to as **segmentation genes.** These genes fall into three general categories: *gap genes, pair-rule genes,* and *segment-polarity genes* (table 16.2; fig. 16.14). These genes are activated sequentially, each by the genes activated before it; each group controls a smaller and more focused domain of the fly's development. In this discussion, we will concentrate on the anterior-posterior system.

The maternal-effect genes of the anterior-posterior system have created Bicoid and Nanos gradients. The segmentation genes increment, narrow, and focus these gradient signals until fourteen distinct bands form, corresponding to the fourteen parasegments that develop, creating compartments that the tissues of the fly arise from (e.g., wings, legs, bristles).

The gap genes were first discovered as mutants that resulted in missing segments in the embryo (fig. 16.14). The Bicoid and Nanos gradients act on gap genes, specifically *hunchback.* Although the Hunchback protein is present in the egg from maternal production, the maternally supplied quantity is apparently not significant. Bicoid and Nanos independently create a Hunchback gradient that is maximal at the anterior end of the embryo, due to activation by Bicoid, and absent at the posterior end, due to Nanos repression. Bicoid is a specific transcription factor that can bind to at least six sites in the promoter region of the *hunchback* gene. Three of these sites are strong binding sites and three are weak. Thus, depending on the concentration of Bicoid in the gradient, different levels of Hunchback are produced, creating the Hunchback gradient. Experiments with extra copies of the *bicoid* gene show that it is the actual quantity of Bicoid present at a particular point, and not the shape of the gradient, that actually determines the effect.

**Table 16.2** Segmentation Genes in *Drosophila*

| Class | Locus | Allelic Designation | Chromosome |
|---|---|---|---|
| Gap | *Krüppel* | *Kr* | 2 |
| | *knirps* | *kni* | 3 |
| | *hunchback* | *hb* | 3 |
| Pair-rule | *paired* | *prd* | 2 |
| | *even-skipped* | *eve* | 2 |
| | *odd-skipped* | *odd* | 2 |
| | *barrel** | *brr* | 3 |
| | *runt* | *run* | 1 |
| | *engrailed* | *en* | 2 |
| Segment-polarity | *cubitus interruptus* | *ci* | 4 |
| | *wingless* | *wg* | 2 |
| | *gooseberry* | *gsb* | 2 |
| | *hedgehog* | *hh* | 3 |
| | *fused* | *fu* | 1 |
| | *patch* | *pat* | 2 |

*Source:* Reprinted with permission from *Nature,* Vol. 287, C. Nusslein-Volhard and E. Wieschaus. Copyright © 1980 Macmillan Magazines Limited.

* *barrel* is a synonym of the *hairy* gene.

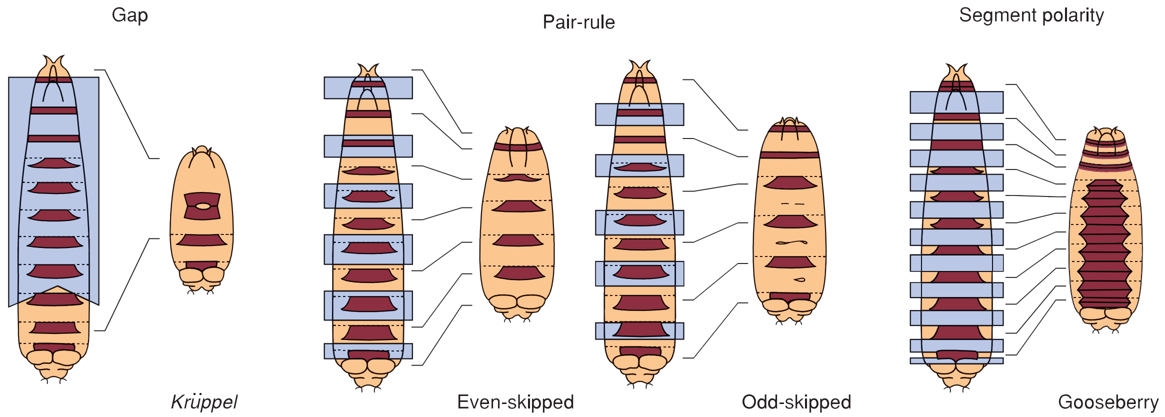


**Figure 16.14** Segmentation genes of *Drosophila* fall into three categories: gap, pair-rule, and segment polarity. To the *left* of each pair is the wild-type larva with cuticular pattern that indicates segment position; to the *right* is the mutant larva. An example of a gap mutation is *Krüppel,* which eliminates the three thoracic and five of the eight abdominal segments (shaded in the wild-type larva). Pair-rule genes are shown that eliminate even (*even-skipped*) or odd (*odd-skipped*) segments (counting from the abdominal segments). An example of a segment polarity gene is *gooseberry,* in which the posterior portion of each segment behaves like a mirror image of the anterior portion of the segment. (Reprinted with permission from Christiane Nüsslein-Volhard and Eric Wieschaus, "Mutations affecting segment number and polarity in *Drosophila,*" *Nature,* 287:795–80, 1980. Copyright © 1980 Macmillan Magazines, Ltd., London, England.)

Presumably, as more Bicoid is present, it binds to more of the *hunchback* promoter sites, resulting in greater transcriptional activity.

At least three gap genes are controlled by the concentrations of the specific transcription factor hunchback: *Krüppel, knirps,* and *giant.* In response to the Hunchback gradient, these three genes are expressed in discrete stripes in the embryo (fig. 16.15). Both anterior and posterior edges of the Krüppel stripe are controlled by Hunchback concentration; Hunchback concentration also controls the anterior edges of the Knirps and Giant stripes. The posterior edges of the Knirps and Giant stripes are controlled by the gradient of the Tailless protein, which is controlled in turn by the terminal maternal-effect gene, *torso* (fig. 16.15). We know the distributions of these proteins by antibody studies, and we know the limits of the protein distributions from studies of various mutants that lack the clear edges of the stripes. For example, the borders of the Krüppel stripe are changed in *hunchback* mutants in accordance with the number of copies of the genes. We have thus gone from very broad and fuzzy regions of maternal-effect gene products to more defined bands of gap gene products.

Interaction of the gap gene proteins then controls transcription of the pair-rule genes (see fig. 16.14). These genes affect alternate sets of segments, even and odd. For example, mutants of the *even-skipped* gene cause the loss of the even-numbered segments, counting by the abdominal segments (loss of two thoracic segments as well

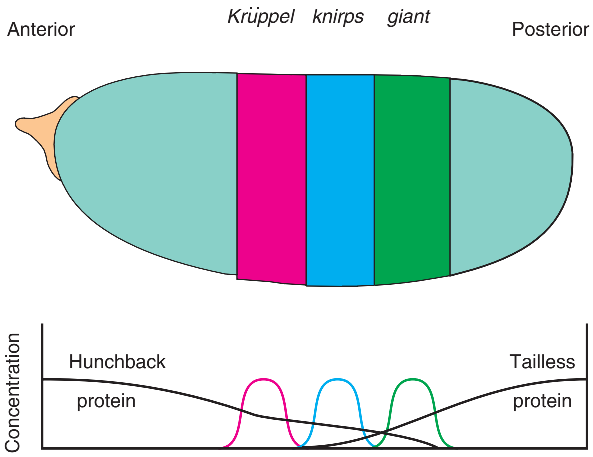


**Figure 16.15** Three discrete bands of gene expression (*Krüppel, knirps,* and *giant*) in the developing *Drosophila* embryo. These bands come about because of the gradients of Hunchback and Tailless proteins. The Hunchback protein level controls the anterior edge of gene expression of *Krüppel, knirps,* and *giant,* as well as the posterior edge of the *Krüppel* gene expression. The Tailless protein level controls the posterior end of *knirps* and *giant* gene expression. The nature of these border edges is verified in mutations of the *hunchback* and *tailless* genes that result in different limits. The three genes (*Krüppel, knirps,* and *giant*) are transcription factors, further controlling gene expression in these regions of the embryo.

as abdominal segments 2, 4, 6, and 8). Finally, the segment-polarity genes are controlled by the pair-rule genes, resulting in genes that affect all segments (see fig. 16.14). For example, mutants of the *gooseberry* gene modify the posterior half of each segment, making it the mirror image of the anterior half.

As development continues, and different classes of segmentation genes are activated, the borders of stripes of activation for these various genes become sharper and sharper, until cell-cell interactions focus the expression of different genes to neighboring cells. For example, we see in figure 16.16 the narrowing and sharpening of the *even-skipped* and *fushi tarazu* bands in the developing embryo. (The gene *fushi tarazu,* meaning "not enough segments" in Japanese, is a pair-rule gene.)

Most segmentation genes are specific transcription factors, genes that interact with DNA to activate or repress transcription. Thus, pattern formation in development is a process of activating different genes in sequence, gradually narrowing the scope of which cells express a particular gene. There is one final group of genes we will discuss in this developmental cascade in *Drosophila.* At this early stage of development, these genes, the **homeotic genes,** take control of the development of the segments.

### Homeotic Mutants

In homeotic mutants, one cell type follows the developmental pathway other cell types normally follow. These genes define the future development of segments based on the pattern of expression of the segmentation genes before them. When they mutate, they switch the development of that segment to an adjacent segment, usually anterior to it. Homeotic genes are also called *memory genes* because they set the developmental fate of a segment, a fate that is

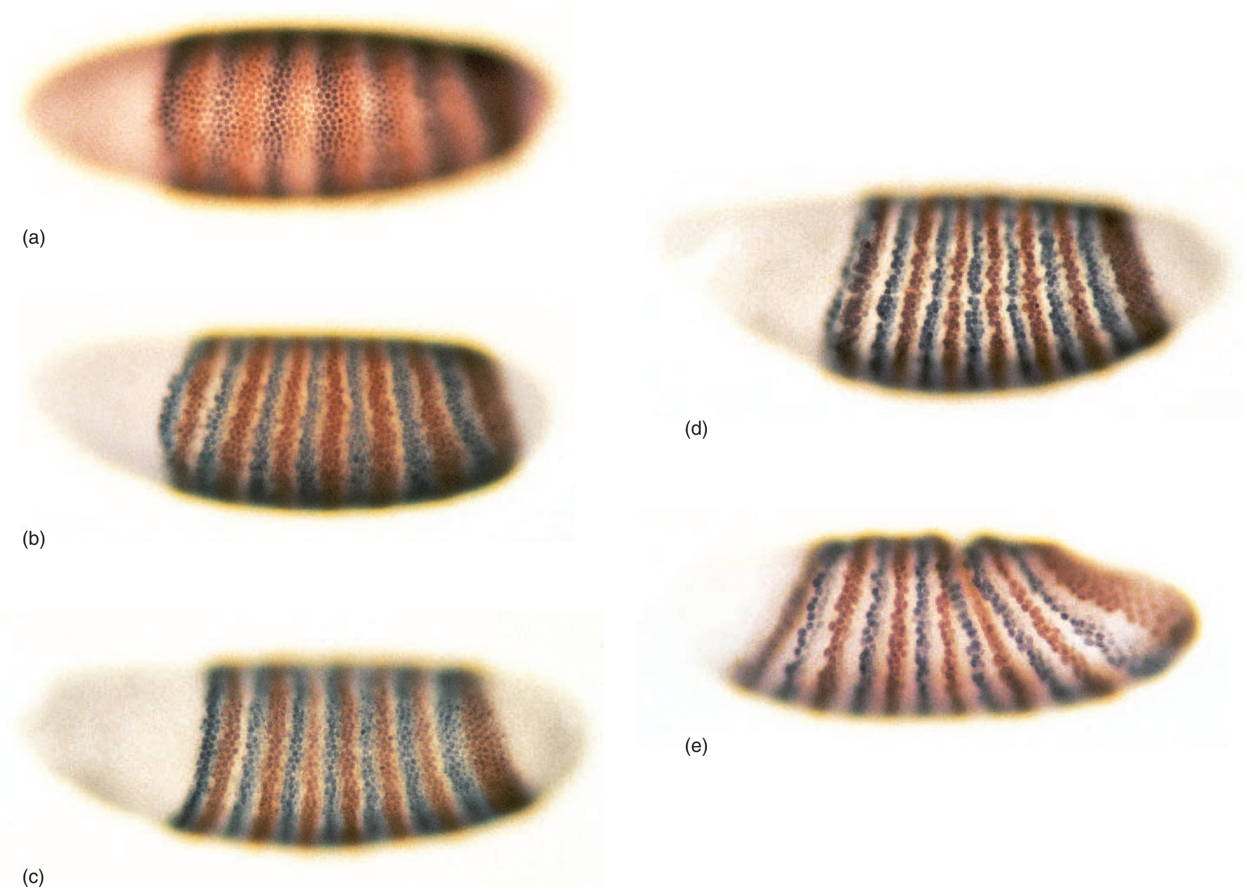


**Figure 16.16** Photos *a–e* show how the margins of expression of two gap genes, *fushi tarazu (brown)* and *even-skipped (gray),* narrow and sharpen as time goes on (between about hours 3 and 4 of embryonic development). The stripes appear from staining with antibodies against the proteins. (From P. A. Lawrence, "The making of a fly," Blackwell Publications, 1992.)

"remembered" from one cell division to the next. They are also called *master-switch* genes since they control the activity of many other genes.

Two major homeotic gene complexes are known in *Drosophila melanogaster* (fig. 16.17): the *bithorax* complex (*BX-C*), analyzed extensively by E. Lewis, D. Hogness, and their colleagues, and the *Antennapedia complex* (*ANT-C*), worked on extensively by W. Gehring, T. Kaufman, and their colleagues. (The two regions together are known as the *Hom-C region.*) Genes in the *Antennapedia* complex control the fate of the anterior development of the fruit fly (head and anterior thorax), whereas genes in the *bithorax* complex control the fate of posterior development (posterior thorax and abdomen). Mutations in the genes of these complexes can change the fate of development of whole sections of the fly. For example, *Nasobemia,* an *Antennapedia*-complex mutant, causes legs to grow where antennae would normally be located (fig. 16.18); and *bithorax,* a *bithorax*-complex mutant, produces flies with two thoraxes (four-winged diptera; fig. 16.19). The genes in these complexes are arranged in order of their progressive action from anterior to posterior on the fly (see fig. 16.17). One model of action for these genes suggests that they require the action of the genes of the adjacent anterior segment plus the action of that homeotic gene itself. Thus, loss of function of a particular gene by mutation would cause a segment to develop like the previous section in the anterior direction.

Edward B. Lewis (1918– ). (Courtesy of Dr. Edward B. Lewis.)

Walter J. Gehring (1939– ). (Courtesy Dr. Walter J. Gehring.)

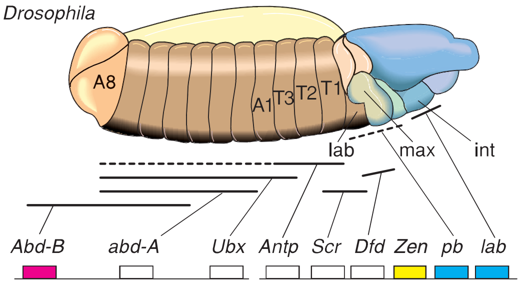


**Figure 16.17** A map of the homeotic complexes *ANT-C* and *BX-C* in *Drosophila,* and the regions of the body in which the genes are expressed, mapped on a ten-hour embryo. Note that the genes are expressed from *right* to *left,* or in an anterior to posterior direction. *Dotted lines* indicate lack of detectable function at this stage in development. Embryonic segments are intercalary (int), maxillary (max), labial (lab), thoracic (T1–T3), and abdominal (A1–A8). Genes are *labial (lab)*, *proboscipedia (pb), Zerknüllt (Zen), Deformed (Dfd), Sex combs reduced (Scr),* and *Antennapedia (Antp)* in *ANT-C* and *Ultrabithorax (Ubx), abdominal-A (abd-A),* and *Abdominal-B (Abd-B)* in *BX-C.* Note that *Zen* is unique in specifying information for the dorsal-ventral axis rather than the anterior-posterior axis. (Reprinted from *Cell,* Vol. 68, W. McGinnis and R. Krumlauf, pp. 283–302, Copyright © 1992, with permission from Elsevier Science.)

### *The Homeo Box*

Using recombinant DNA techniques, W. Gehring and his colleagues found a consensus sequence of 180 base pairs of DNA in genes of the *Antennapedia* and *bithorax* complexes. Further probing localized this same segment of 180 base pairs to about a dozen genes in *Drosophila,* all with homeotic or segmentation properties. They thus called this DNA sequence the **homeo box.** The nucleotides of the homeo box are translated into a peptide region of 60 amino acids called the **homeo domain** (fig. 16.20, box 16.1).

Using a recombinant probe for the homeo box, or a computer search for the consensus sequence, researchers found it in the genes of plants, yeast, sea urchins, frogs, and human beings. This high degree of se-

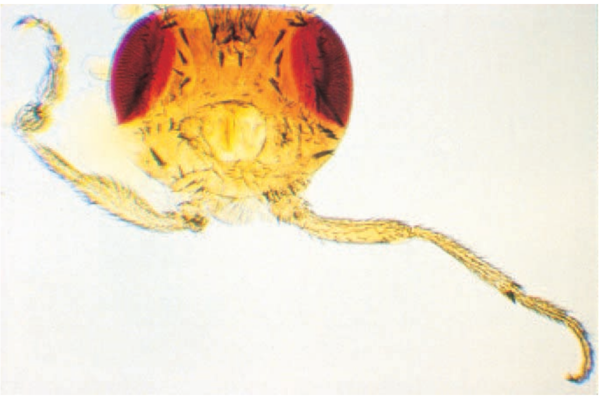


**Figure 16.18** *Nasobemia,* a mutation that causes legs to grow in the place of antennae on the head of a *Drosophila.* (Courtesy of Dr. Walter J. Gehring.)

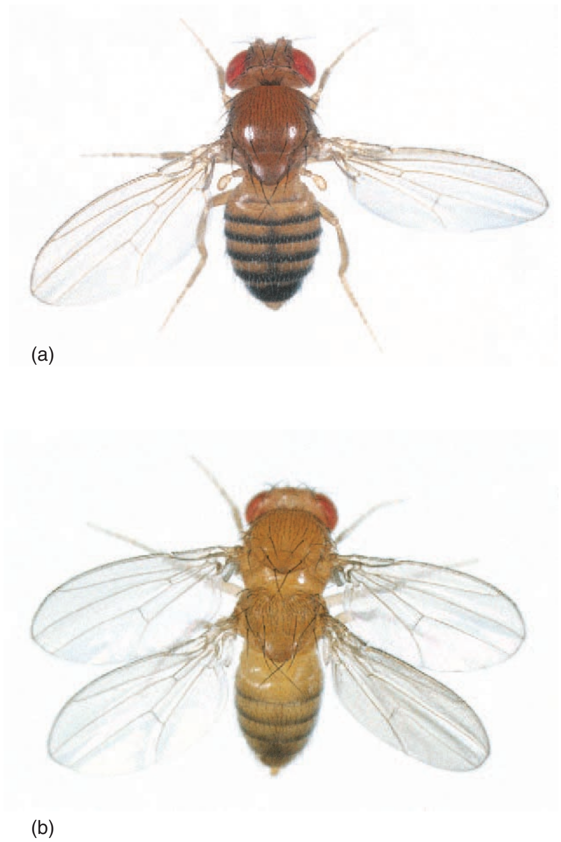


**Figure 16.19** A normal fruit fly (*a*) and a *bithorax* mutant (*b*). The *bithorax* mutant is actually the product of a combination of three mutations that produce a fly with an almost perfect second thorax with its own set of wings. (Courtesy of E. B. Lewis, California Institute of Technology.)

quence conservation across widely divergent groups of organisms indicates that the sequence is crucial to the functioning of homeotic genes and that the mechanism arose early in evolutionary time.

The conservation of homeotic control throughout evolutionary history is evident from the fact that the homeotic genes that control the development of fruit flies are also found in mammals. During evolutionary history, the mammalian genome has been duplicated four times. Thus, there are four homeotic clusters in mice, called *Hox clusters,* on four different chromosomes (fig. 16.21). With multiple copies, the genes could be modified by evolution while still maintaining one copy functioning as originally intended. This duplication has allowed increased complexity in higher eukaryotes. As the *Drosophila* homeobox genes function from anterior to posterior, so do the homeobox genes in other organisms.

## Plants

Much work is being done in determining the genetic control of development in plants. A favored model is the thale cress, *Arabidopsis thaliana,* a member of the mustard family (fig. 16.22). It is a dicotyledonous angiosperm, ideal for the study of flower development, a current focus of attention. Flowers have an arrangement of repeated units not unlike the segmentation found in fruit flies.

Flower development takes place in two phases, *floral induction* and *pattern formation.* In floral induction, the **shoot apical meristem** sets aside a **floral meristem.** The organ primordia are then generated. There are four primordia, in the form of four whorls, that make up a flower. Outermost is the *sepal whorl,* then the *petal whorl,* then the *stamen whorl,* responsible for the male parts of the flower, and finally the innermost *carpel whorl,* responsible for the female parts of the flower (the pistil; fig. 16.23). The genetics of development in plants is

| | 1 | | | | | | | | | | | | | | | | | | | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Antennapedia* | Arg | Lys | Arg | Gly | Arg | Gln | Thr | Tyr | Thr | Arg | Tyr | Gln | Thr | Leu | Glu | Leu | Glu | Lys | Glu | Phe |
| *Mouse MO-10* | Ser | Lys | Arg | Gly | Arg | Thr | Ala | Tyr | Thr | Arg | Pro | Gln | Leu | Val | Glu | Leu | Glu | Lys | Glu | Phe |
| *Frog MM3* | Arg | Lys | Arg | Gly | Arg | Gln | Thr | Tyr | Thr | Arg | Tyr | Gln | Thr | Leu | Glu | Leu | Glu | Lys | Glu | Phe |

| | 21 | | | | | | | | | | | | | | | | | | | 40 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Antennapedia* | His | Phe | Asn | Arg | Tyr | Leu | Thr | Arg | Arg | Arg | Arg | Ile | Glu | Ile | Ala | His | Ala | Leu | Cys | Leu |
| *Mouse MO-10* | His | Phe | Asn | Arg | Tyr | Leu | Met | Arg | Pro | Arg | Arg | Val | Glu | Met | Ala | Asn | Leu | Leu | Asn | Leu |
| *Frog MM3* | His | Phe | Asn | Arg | Tyr | Leu | Thr | Arg | Arg | Arg | Arg | Ile | Glu | Ile | Ala | His | Val | Leu | Cys | Leu |

| | 41 | | | | | | | | | | | | | | | | | | | 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Antennapedia* | Thr | Glu | Arg | Gln | Ile | Lys | Ile | Trp | Phe | Gln | Asn | Arg | Arg | Met | Lys | Trp | Lys | Lys | Glu | Asn |
| *Mouse MO-10* | Thr | Glu | Arg | Gln | Ile | Lys | Ile | Trp | Phe | Gln | Asn | Arg | Arg | Met | Lys | Tyr | Lys | Lys | Asp | Gln |
| *Frog MM3* | Thr | Glu | Arg | Gln | Ile | Lys | Ile | Trp | Phe | Gln | Asn | Arg | Arg | Met | Lys | Trp | Lys | Lys | Glu | Asn |

**Figure 16.20** The homeo domain of three genes: the *MO-10* gene from the mouse *(Mus),* the *MM3* gene from the frog *(Rana),* and the *Antennapedia* gene from *Drosophila,* which is considered the consensus sequence; amino acids in *red* differ from this sequence. (From W. J. Gehring, *Scientific American,* November 1985. Reprinted with permission of Walter J. Gehring.)

## BOX 16.1

## Experimental Methods

### *Protein Motifs of DNA Recognition*

Different motifs have been found in specific transcription factors and other proteins that bind to DNA. In homeo domains, amino acids 31 to 38 and 41 to 50 form α helices. The configuration of two α helices in a protein, separated by a short segment (called a "turn"), has been found in many proteins that bind to DNA (e.g., Cro, λ repressor, CAP protein). It is called the **helix-turn-helix motif.** One α helix recognizes a DNA sequence by fitting into the major groove, and the other helix stabilizes the configuration (fig. 1).

The helix-turn-helix (or helix-loop-helix) motif appears in some proteins that bind to DNA. However, different motifs have also been found in other proteins that bind to DNA. These include the **zinc finger,** the **leucine zipper,** and the **basic/helix-loop-helix/leucine zipper.** The zinc finger, a fingerlike projection of amino acids, whose base consists of cysteine and histidine residues binding a zinc ion, was first discovered in 1985 by A. Klug and his colleagues in the transcription factor TFIIIA in *Xenopus* (fig. 2). These fingers are referred to as $C_2H_2$ proteins because two cysteines ($C_2$) and two histidines ($H_2$) are involved. There are also $C_x$ proteins in which $x$ is either 4, 5, or 6, referring to the number of cysteines involved in the chelation of the zinc ion, and other variants of protein structures formed around zinc ions.

Another motif was discovered in analyzing a DNA-binding protein from rat liver nuclei. Scientists noticed that in α-helical regions of the protein, a repetition of leucines occurred every seven residues for sequences as long as forty-two residues. In a helical configuration, these leucines would line up on one side of the protein. When a computer search for sequences of this type was done, several other proteins, believed to bind to DNA, showed up with this configuration, including three cancer-causing genes, *c-myc, fos,* and *jun,* and a transcription-regulating protein in yeast. Using the computer, the scientists developed the leucine-zipper model, in which two helices with leucine repeats would interdigitate the leucines, in zipper fashion, to form a stable molecule (fig. 3). This zipper could provide a scaffolding for other amino acids that could then recognize specific DNA sequences in order to perform their functions.

A recently discovered DNA-binding motif, the basic/helix-loop-helix/leucine zipper, is a series of basic amino acids followed by the helix-loop-helix and then a leucine zipper (fig. 4). This motif is found in the Myc oncoprotein and in a transcription factor, Max, that binds with Myc. Knowing that specific motifs bind to DNA gives us an idea of the function of many proteins as soon as their amino acid sequences are determined.

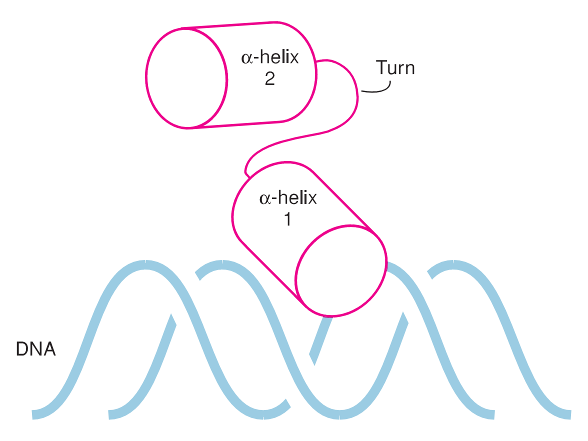


**Figure 1** The helix-turn-helix motif of a DNA-binding protein. The two helices are pictured as cylinders. The α-helix 1 recognizes the DNA sequence in the major groove; the α-helix 2 stabilizes the configuration.

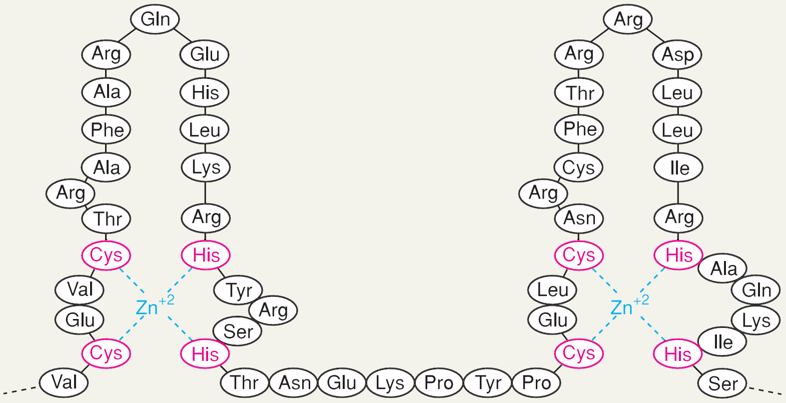

Gln
Arg
Glu
Ala
His
Phe
Leu
Ala
Lys
Arg
Thr
Arg
Cys
His
Val
Tyr
Zn$^{+2}$
Arg
Glu
Ser
Cys
His
Val
Thr
Asn
Glu
Lys
Pro
Tyr
Pro
Arg
Arg
Asp
Thr
Leu
Phe
Leu
Cys
Ile
Arg
Asn
Arg
Cys
His
Ala
Leu
Gln
Zn$^{+2}$
Glu
Lys
Ile
Cys
His
Ser

**Figure 2** The zinc-finger configuration of the TFIIIA protein. Zinc chelates with cysteines and histidines to form the base of the finger structure.

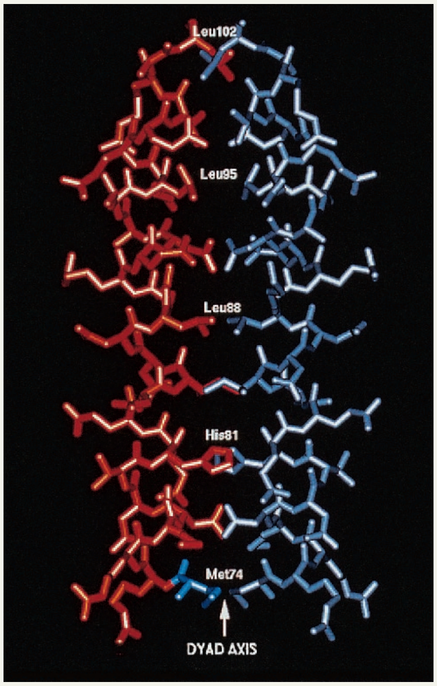

Leu102
Leu95
Leu88
His81
Met74
DYAD AXIS

**Figure 3** Three-dimensional model of the leucine-zipper region of the Max transcription factor. The leucine residues line up opposite each other in the two strands. (From A. R. Ferré-D'Amaré, et al., "Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain," *Nature* 363:38–45, May 6, 1993. © Macmillian Magazines Limited.)

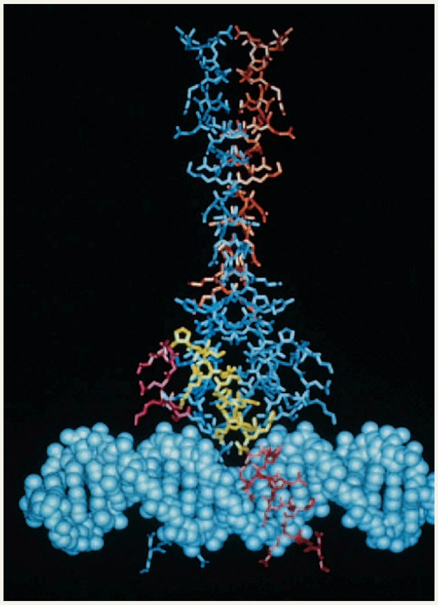


**Figure 4** Diagram of a dimer of basic/helix-loop-helix/leucine zipper interacting with DNA. One basic region, interacting with DNA, is shown in *red,* followed by the first helix in *yellow,* the loop in *purple,* the second helix in *blue,* and the zipper portion in *orange.* The second monomer is shown in *gray.* (From A. R. Ferré-D'Amaré, et al., "Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain," *Nature* 363:38–45, May 6, 1993. © Macmillan Magazines Limited.)

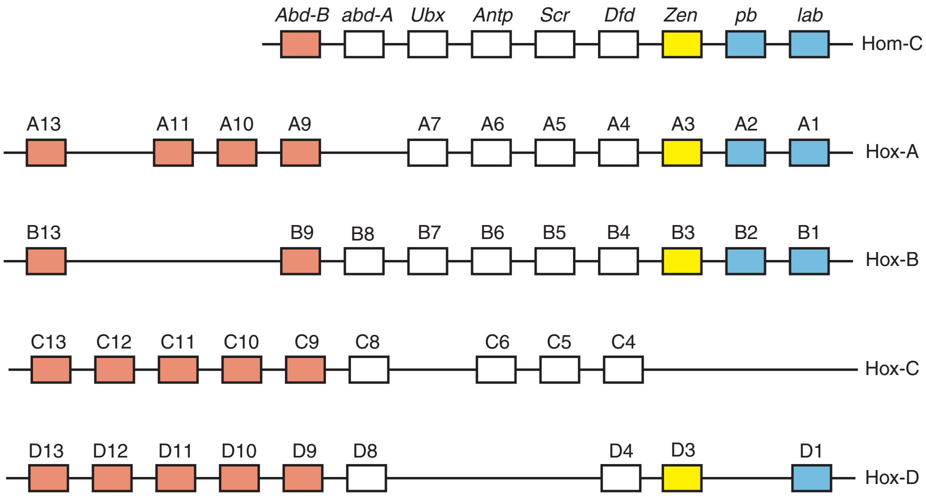

Abd-B
abd-A
Ubx
Antp
Scr
Dfd
Zen
pb
lab
Hom-C
A13
A11
A10
A9
A7
A6
A5
A4
A3
A2
A1
Hox-A
B13
B9
B8
B7
B6
B5
B4
B3
B2
B1
Hox-B
C13
C12
C11
C10
C9
C8
C6
C5
C4
Hox-C
D13
D12
D11
D10
D9
D8
D4
D3
D1
Hox-D

**Figure 16.21** The homeobox genes in *Drosophila* (Hom-C) are aligned with the four homeobox clusters of the mouse, labeled Hox-A, Hox-B, Hox-C, and Hox-D. Note that not all genes are present in all four mouse Hox clusters and that as many as four additional genes (10–13) are present in each mouse region as compared with the fly.

**Figure 16.22** The thale cress plant, *Arabidopsis thaliana.* (Courtesy of Dr. John Celenza.)

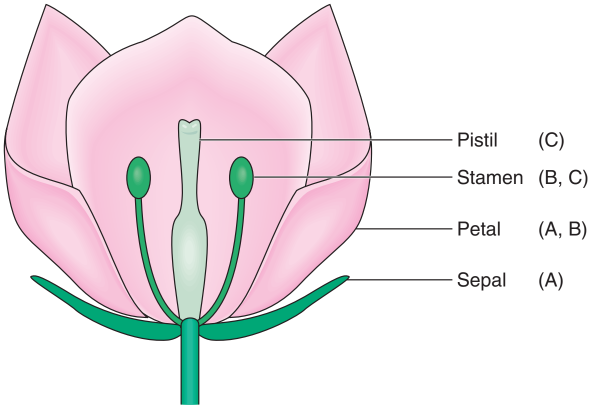

Pistil (C)
Stamen (B, C)
Petal (A, B)
Sepal (A)

**Figure 16.23** Cutaway view of a typical angiosperm flower. The flower develops from four whorls: sepal, petal, stamen, and carpel. Homeodomain genes in the A group are active in sepal and petal whorls; homeodomain genes of the B group are active in petal and stamen whorls; and homeodomain genes in the C group are active in the stamen and carpel whorls (the pistil develops from the carpel whorl).

studied by mutational analysis, selective ablation (removal or killing) of cells during development, and other techniques used in animal studies.

Many genes have been isolated that affect the sequence of steps of floral induction and pattern formation. The first stage to be controlled in floral induction is its timing. That is, flower formation usually occurs at a specific time in the life cycle of a plant, affected by environmental cues (day length, temperature). In *Arabidopsis,* at least three dozen genes have been isolated that affect the timing of flower formation. These genes include

*CONSTANS,* a late-flowering gene, *EARLY FLOWERING 1,* an early-flowering gene, and *GIBBERELLIN INSENSITIVE,* a gene for late flowering only in short days (autumn).

The next stage in floral induction is generating floral meristem at the point where a flower will form. At least five genes are known that impart identity on floral meristem **(floral-meristem identity genes);** when mutated, these genes result in either shoots instead of flowers or in highly abnormal flowers. These genes include *LEAFY, UNUSUAL FLORAL ORGANS, APETALA1,* and *APETALA2.*

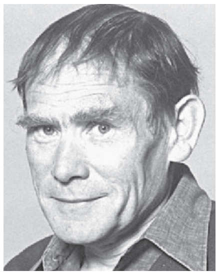
Floral development continues by the creation of organ primordia. Although far removed from animals in both taxonomy and DNA sequences, plants have homeotic genes, some producing proteins homologous to those produced by animal genes. Currently, floral homeotic genes are classified into three categories, A, B, and C. Genes from category A affect sepals and petals; genes from category B affect petals and stamens; and genes from category C affect stamens and carpels (fig. 16.23). This is not unlike the model of action in *Drosophila's* homeotic gene clusters, which acts sequentially, controlling development along the head-to-tail axis of the fly. It appears that genetic control of floral development is highly conserved across angiosperms, the dominant plant group.

An example of a homeotic gene is *AGAMOUS,* a gene in the C group required for the development of stamens and carpels. Expression of this gene takes place in the third and fourth whorls of the flower, the stamen and carpel whorls. After its expression in the appropriate whorls, *AGAMOUS* is repressed. Its repressor is another gene, *CURLY LEAF.* When the protein product of *CURLY LEAF* was compared with protein sequences from *Drosophila,* it proved to have similarities in amino acid sequence with a gene in *Drosophila* called *Enhancer of zeste.* This gene is also a repressor of a homeotic gene, but in fruit flies.

Thus, several valuable conclusions come from this study of *Arabidopsis.* Most important is the fact that plants and animals seem to use similar mechanisms in development. Both groups have repeated units (segments) in development; both have homeotic genes that control developmental pathways in these units; both have repressors of homeotic genes that maintain the proper developmental fate in their segments; and, despite large taxonomic distances, there is some homology between the proteins in plants and animals.

## Other Models of Development

Although the study of development in animals has progressed markedly by using *Drosophila* as a model, other organisms have been used as well. Historically, amphibians were the focus of developmental research because they have large eggs that can be easily observed and manipulated. The same reasoning made the chick embryo a classical model of development. The nematode *Caenorhabditis elegans* has emerged as another model organism for developmental studies because of its simplicity (fig. 16.24). Each individual consists of only about one thousand cells; its life cycle lasts only 3.5 days; and with only $8 \times 10^7$ base pairs of DNA, it has the smallest genome of any multicellular organism. In 1963, S. Brenner proposed learning the lineage of every cell in the adult. With the efforts of numerous colleagues, that work was completed in about twenty years. From the fertilized egg to the adult, the division and fate of every cell of this nematode worm is known. The worm has been especially useful in studying homeotic mutants and



Sydney Brenner (1927– ).
(Courtesy of Dr. Sydney Brenner.)

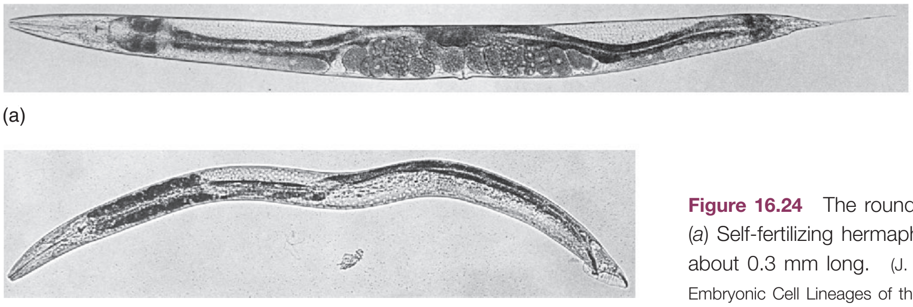


**Figure 16.24** The roundworm *Caenorhabditis elegans.* (*a*) Self-fertilizing hermaphrodite. (*b*) Male. The worms are about 0.3 mm long. (J. E. Sulston and H. R. Horvitz, "Post-Embryonic Cell Lineages of the Nematode *Caenorhabditis elegans,*" *Developmental Biology,* 56:1101–56, 1977, Academic Press.)

**apoptosis,** programmed cell death. As we will see later, apoptosis is important in development as well as in the elimination of infected or cancerous cells. Also being used as animal models are mice and zebra fish. In plants, the snapdragon, *Antirrhinum majus,* is another model organism.

Development is a growth process that is, among other things, an orderly process of cell division. As we discussed in chapter 3, the cell cycle is controlled by checkpoints; the cycle is allowed to continue if the cell is ready for the next stage. "Ready" means, among other things, that the cell has successfully completed DNA replication and repair of DNA damage. If the cell is not ready, the cell cycle stops until the cell is ready. If the cell is damaged beyond repair, including being cancerous, programmed cell death (apoptosis) is initiated. It is clear that numerous checks and balances are involved in assuring that only healthy, ready cells continue in the cell cycle. Interference to these checks and balances can lead to uncontrolled cell growth—cancer.

## CANCER

**Cancer** is an informal term for a diverse class of more than 100 distinct diseases marked by abnormal cell proliferation; white blood cells proliferate at an inappropriate rate, or other cell types form growths known as **tumors (neoplasms).** Benign tumors grow in only one place and do not invade other tissues. The cells of malignant tumors not only continue to proliferate but also invade nearby tissues and, by a process called **metastasis,** spread to distant parts of the body through blood or lymph vessels and start new centers of uncontrolled cell growth wherever they go.

Cancers are generally divided into four groups, dependent on the type of cells originally involved. Two types of cancer cause overproduction of white blood cells. **Leukemias** are diseases that cause excessive production of leukocytes, which originate in the bone marrow. **Lymphomas** cause excessive production of lymphocytes, which originate in the lymph nodes and spleen. **Sarcomas** are tumors of tissue such as muscle, bone, and cartilage that arise from the embryological mesoderm. About 85% of cancers are **carcinomas,** tumors arising from epithelial tissue such as glands, breast, skin, and the linings of the urogenital, digestive, and respiratory systems.

All cancers are genetic: They come about from alterations in genes that control cell growth. Most evidence indicates that cancers are *clonal*—they arise from a single aberrant cell that then proliferates. Therefore, analyzing the causes of cancer comes down to trying to understand how one cell is changed, or *transformed,* from a normal cell to a cancerous one. As we will see, most cancers come about from a series of genetic changes, progressing from an aberrant cell to an aggressively cancerous one. This view is called the **clonal evolution theory** of cancer.

Historically, cancers were understood to be caused by either mutation or by viruses. We now know that viruses can bring cancer-causing genes into cells, where their mutated form or inappropriate location can lead to cancer. Thus, both the mutational and viral views of cancer are ultimately concerned with mutation. In essence, cancers result from the inappropriate activity of certain genes, whether those genes were changed by mutation or were imported or activated by viruses.

### Mutational Nature of Cancer

Mutations, both point and chromosomal, have been implicated in carcinogenesis (table 16.3). For example, the disease **xeroderma pigmentosum** in human beings is caused by mutations in any of seven loci (*XpA-XpG*) that inactivate the mutation repair system that corrects UV-light damage (see chapter 12); exposure to the sun then results in skin lesions that often become malignant. A related disease, **ataxia-telangiectasia,** is caused by a defect in the double-strand break repair mechanism, often the result of X-ray induced damage. (*Ataxia* refers to difficulty in balance; *telangiectasia* refers to dilated blood vessels in the eye membranes.) By binding to the ends of DNA, as would happen when a double-strand break occurs, ATM (the protein product of the *ataxia-telangiectasia* locus, *atm*) begins a signaling pathway that tells the cell there are broken ends of DNA. Persons with this defect are at risk for acute and chronic leukemia and lymphomas; women with it are also at risk for ovarian cancers.

Most cancers are associated with chromosomal defects; improved chromosomal banding techniques have demonstrated that a specific chromosomal defect is often associated with a specific cancer (table 16.3, fig. 16.25, box 16.2). The implication is that when a gene is in a new location (because of translocation or the deletion of intervening material), that gene may fall under the control of more powerful promoters or promoters outside the range of that gene's normal control. As we shall see, genes that are known to be able to transform cells **(oncogenes)** are often the ones that are relocated into regions of new control. These oncogenes then become more active, and transformation follows.

#### Cancer-Family Syndromes

In some cases, a predisposition for malignancies is inherited. When four thousand clinic registrants were interviewed, almost half reported virtually no family history of

**Table 16.3** **A Small Sample of Chromosomal Rearrangements Associated with Specific Cancers**

| Disease | Chromosomal Rearrangement | Genes Affected |
|---|---|---|
| Burkitt's lymphoma | t(8; 14)* | *c-MYC* |
| Non-Hodgkin's lymphoma | t(3; 4) | *Laz3, BCL-6* |
| B-cell chronic lymphocytic leukemia | t(11; 14) | *BCL-1, PRAD-1* |
| Follicular lymphoma | t(14; 18) | *BCL-2* |
| T/B-cell lymphoma | Inversion, chromosome 14 | *TCR-a* |
| Chronic myelogenous leukemia/ Acute lymphocytic leukemia | t(9; 22) | *CABL* |
| Ewing's sarcoma | t(11; 22) | *FLI1, EWS* |
| Melanoma of soft parts | t(12; 22) | *ATF1, EWS* |
| Liposarcoma | t(12; 16) | *CHOP, FUS* |

Reprinted with permission from *Nature,* Vol. 372, T. H. Rabbits. Copyright © 1994 Macmillan Magazines Limited.

* The notation of the form t(8; 14) indicates a *t*ranslocation between chromosomes *8* and *14.*

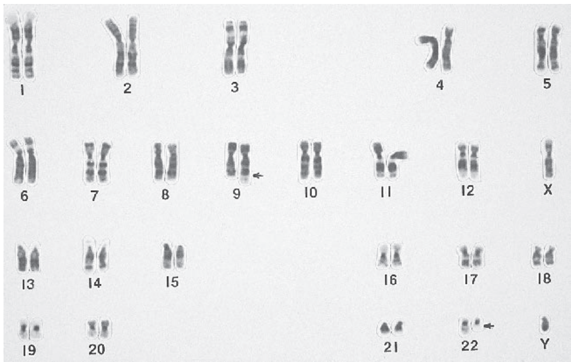


**Figure 16.25** G-banded chromosomes from a patient with chronic myelogenous leukemia showing a translocation of chromatin *(arrows)* from chromosome 22 to chromosome 9. (Courtesy of Charles Rubin, M.D., University of Chicago, Department of Pediatric Hematology/Oncology.)

cancer, whereas about 7% reported that many family members had cancer. This 7% was considered cancer prone because three or more close relatives of the interviewed person had cancer. The interpretation of the study is that some families are predisposed toward cancer, but most are not, displacing the idea that everyone in the population has a uniform and low probability of developing cancer. Lending support to this interpretation are the **cancer-family syndromes,** in which family members seem to inherit a nonspecific predisposition toward tumors of various types. At least twenty cancer-family syndromes are known. In figure 16.26, we see a pedigree for a cancer-family syndrome in which the predisposition for several different types of cancers, rather than a particular type of cancer, seems to be inherited. Women in this family get breast, colon, and pancreatic cancers, whereas men get colon, prostate, and pancreatic cancers.

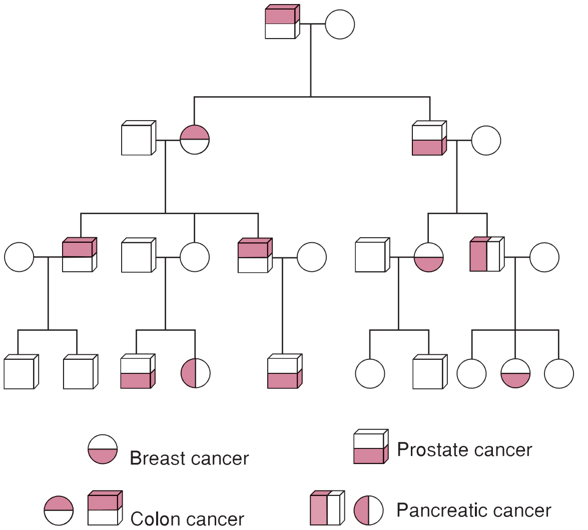


**Figure 16.26** Pedigree of a type I cancer-family syndrome. This is interpreted as the inheritance of the propensity toward cancer rather than the inheritance of any specific type of cancer.

### *Tumor-Suppressor Genes*

There is a class of cancer-related genes called the **tumor-suppressor genes** (also called **anti-oncogenes**). These genes act by suppressing malignant growth. Mutations are recessive, and in the homozygous state, cancer

**BOX 16.2**

## Experimental Methods

### *Chromosomal Painting*

A technique has been developed that allows investigators to differentiate all of our chromosomes very quickly and accurately by seeing them painted in different fluorescent colors. This technique allows a scientist or clinician to determine quickly whether any chromosomal anomalies exist, either in number (aneuploidy) or structure (deletions, translocations). The technique, **chromosomal painting,** is a variant of the technique known as **fluorescent in situ hybridization (FISH),** in which a fluorescent dye is attached to a nucleotide probe that then binds to a specific site on a chromosome and makes itself visible by fluorescence (see fig. 13.41). A whole chromosome can be made visible by this technique if enough probes are available to mark enough of the chromosome. However, there are not enough fluorescent markers known to paint all 24 of our chromosomes (autosomes 1–22, X, Y) a different color. Now, with as few as five different fluorescent markers and enough probes to coat each chromosome, it is possible to make combinations of the different marker dyes so that each chromosome fluoresces a different color. Because the colors are not generally distinguishable by the human eye, they have to be separated by a computer that then assigns each chromosome its own color. As figure 1 shows, the technique works very well. With it, we can rapidly determine any chromosomal anomaly in a given cell. This technique is helpful in clinical diagnosis of various syndromes and diseases, including cancer.

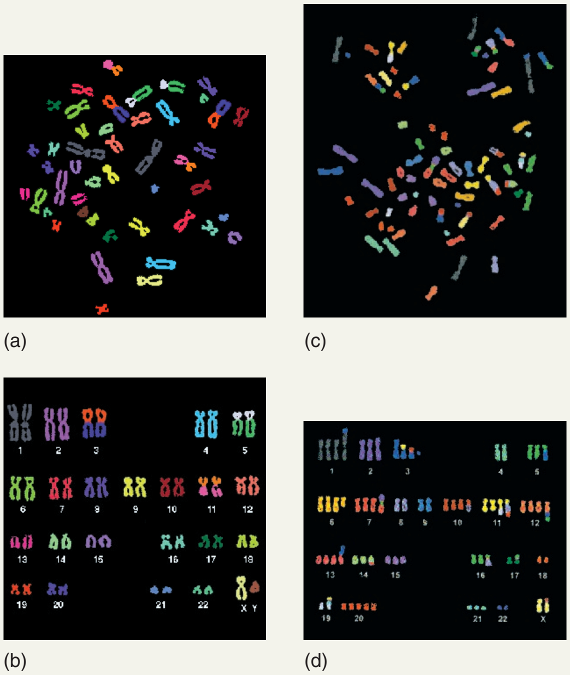


**Figure 1** Chromosomal spreads after treating with probes specific for all human chromosomes and attached to flourescent tags. Colors are generated by computer. *Left (a, b)* are the spread and karyotype of a normal cell; *right (c, d)* are the same for an ovarian cancer cell with complex chromosomal anomalies. (Courtesy of Michael R. Speicher and David C. Ward, "The coloring of cytogenetics," *Nature Genetics,* 2:1046–48, 1996, figs. 2 and 3. Photos courtesy David C. Ward.)

ensues. The first tumor-suppressor gene to be isolated was the gene for **retinoblastoma,** a tumor of retinoblast cells, which are precursors to cone cells in the retina of the eye. This is a disease young children contract, because after the retinoblast cells differentiate, they no longer divide and apparently can no longer form tumors. The disease occurs both in a hereditary and a sporadic form. Both forms are presumably due to the recessive homozygous state of the locus. In the hereditary form, individuals inherit one mutant allele; a second mutation results in the disease. In the sporadic form, with identical symptoms, both alleles have apparently mutated spontaneously in the somatic tissue of the retina. The retinoblastoma gene has also been implicated in other cancers, including sarcomas and carcinomas of the lung, bladder, and breast.

How do we know that retinoblastoma results from the loss of suppression rather than simply the activity of an oncogene? J. Yunis, who examined cells from several retinoblastoma patients, found a frequently deleted part of chromosome 13, specifically band q14. Yunis noticed that the exact points of deletion varied from individual to individual, indicating that the phenomenon was due to loss of gene action rather than enhancement of gene activity due to the new placement of genes previously separated by the deleted material.

Under normal circumstances, the retinoblastoma protein, RB, inhibits the cell cycle from advancing. If the appropriate checkpoint is passed, RB is phosphorylated by cyclin-dependent kinase and cyclin complexes, and the cell cycle progresses. As the protein product of a homozygous recessive mutant, RB no longer inhibits the cell cycle advance, even if the checkpoint has not been cleared. Thus, DNA-damaged and cancerous cells are allowed to continue to grow.

The retinoblastoma gene has been isolated and cloned. The gene specifies a 105-kilodalton protein (p105) found in the nucleus, as would be expected if it were a suppressor of DNA transcription. It binds with at least three known oncogenic proteins: the E1A protein of adenovirus, the SV40 (a simian virus) large T antigen, and the 16E7 protein of human papillomavirus, a virus associated with 50% of cervical carcinomas. The implications are that these three viruses may use a similar mechanism in transformation, and this mechanism involves inactivation of the retinoblastoma p105 protein.

Further support for the existence of tumor-suppressor genes came from work by E. Stanbridge and his colleagues with another childhood cancer, **Wilm's tumor.** This is a kidney cancer that is also believed to be caused by loss of action in a tumor-suppressor gene. It is associated with the loss of band p13 on chromosome 11. Researchers introduced a normal chromosome 11 into Wilm's tumor cells growing in culture. The result was normal cell growth, exactly what we would predict if the introduced normal gene were a tumor-suppressor gene.

A third tumor-suppressor gene is the *p53* gene, named for its 53-kilodalton protein product and located on chromosome 17. This gene is the most common mutation in cancers, found in more than 50% of human tumors. It achieved the status of *Science* magazine's 1993 "Molecule of the Year." Since the *p53* protein is found in so many cases, it is clear that its role as a tumor suppressor was of great importance in the normal activity of cells. Normally, p53 is highly unstable: the MDM2 protein binds its amino terminal end and ubiquinates it, leading to the rapid degradation of p53 in the proteasome within several minutes. However, p53 is stabilized when it is phosphorylated by cell-cycle checkpoint kinases. For example, ATM binds to double-stranded DNA breaks. Bound this way, it activates the protein CHK2, a checkpoint kinase that then phosphorylates p53. In the active state, p53 is a transcription factor that induces at least thirty-four different genes, genes involved either in stopping the cell cycle, inducing apoptosis, or regulating itself.

First, p53 stops the cell cycle to give the cell a chance to repair its DNA. Cell growth is arrested by the induction (also called *upregulation*) of cyclin-dependent kinase inhibitors (proteins such as p21, WAF1, and CIP1). This action stops the cell cycle. In fact, if DNA repair does not take place, cells can be forced to remain permanently in G1 phase. Alternatively, p53 can induce cell death by upregulating the *bax* gene. Its protein is involved in the pathway to induce *caspases,* proteinases that destroy the cell. (Caspases get their name from the fact that they are *c*ysteine-requiring *asp*artic acid protein*ases.* The *bax* gene's name comes from b*cl-2 a*ssociated-*x* gene; *bcl-2* is from *B-c*ell *l*eukemia/*l*ymphoma-*2.*)

Finally, the *p53* protein is a transcription factor for the gene for MDM2, the protein that regulates *p53.* Thus, the *p53* protein has a narrow window in which to stop the cell cycle or induce apoptosis, giving the cell a chance to repair its DNA damage or commit suicide. After this, the *p53* protein is itself repressed (fig. 16.27).

It is clear that the loss of *p53* activity allows DNA damage to build up in a cell. This is why more than 50% of cancers involve loss of *p53* activity. More than twenty other tumor-suppressor genes are known.

## Viral Nature of Cancer

### Retroviruses

Animal viruses come in many different varieties, with DNA or RNA as their genetic material (fig. 16.28). Several classes of viruses, both DNA and RNA, can transform cells, a process that may or may not be caused by an oncogene the virus carries. Some DNA viruses do carry oncogenes, such as the adenovirus that carries the gene for the E1A protein, which may act by binding to the retinoblastoma repressor protein. Oncogenes, however, were originally discovered in retroviruses, a group of very simple RNA viruses that contain the enzyme reverse transcriptase. After the virus enters the host cell, this enzyme converts the viral RNA into DNA. In 1910, Peyton Rous, who much later won the Nobel Prize for his work, discovered that a sarcoma in chickens could be induced by a cell-free extract from a tumor in another chicken. The transmitted agent was later found to be a retrovirus, named Rous sarcoma virus, the first retrovirus to be discovered.

Peyton Rous (1879–1970). (Courtesy of Rockefeller University Archives.)

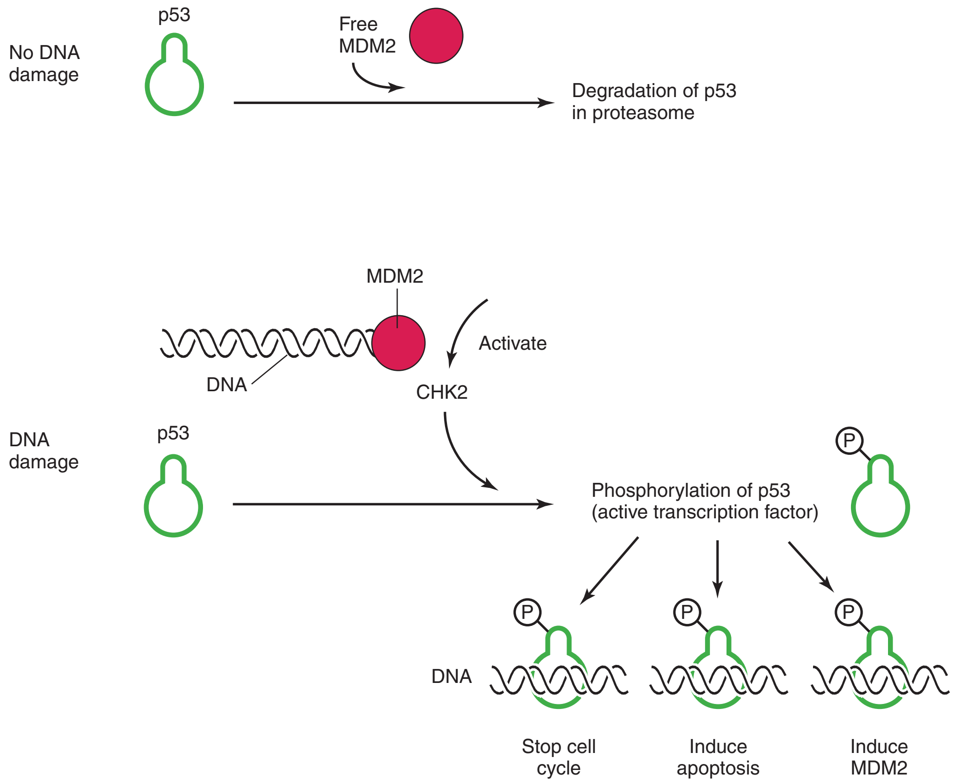


**Figure 16.27** Normally, MDM2 ubiquinates the p53 protein, which leads to its degradation in the proteasome. If there is DNA damage that produces broken DNA ends (double-strand breaks), MDM2 binds to these ends and activates the checkpoint kinase CHK2. CHK2 phosphorylates p53, making it stable and an active transcription factor. Next, the active p53 binds to the promoters of synexpression group genes for stopping the cell cycle and apoptosis. In addition, p53 induces the *MDM2* gene, which can then cause p53 degradation if the broken DNA has been repaired. (MDM2 is named for the cell culture in which it was cloned—*m*urine *d*ouble *m*inute chromosome clone number *2*.)

The retrovirus, which often carries only three genes, integrates into the host genome in a series of steps (fig. 16.29). When the virus enters the host, it is in the form of a plus (+) RNA strand (capable of acting as a messenger RNA; the minus strand is the complement to the strand). At either end is a repeated sequence (R) located outside two unique sequences (U3 and U5). Through reverse transcription, using the reverse transcriptase the virus brings in, the viral RNA is converted to a double-stranded DNA. During that process, the ends of the DNA take on the configuration of long terminal repeats (LTRs), repetitions of U3-R-U5. The linear DNA then circularizes and integrates into the host genome just as a transposon does, generating short direct repeats at either end.

As we mentioned, retroviruses can cause cellular transformation through direct integration or from the oncogenes they carry. Transformation from integration comes about because the integrated provirus either inactivates a tumor-suppressor gene or activates an oncogene in a process called **insertion mutagenesis.** The U3 region of the retrovirus contains both an enhancer and a promoter. Since a long terminal repeat lies at either end of the provirus, cellular genes can be turned on when the virus integrates.

### Oncogenes

Genetic analysis and recombinant DNA studies showed that Rous sarcoma virus transforms cells through the action of a single gene. This gene, called *src* for sarcoma, was the first viral oncogene discovered. Since then at least fifty have been discovered, and each has been given a three-letter designation (table 16.4). Unlike tumor suppressors, which lead to cancer when in the homozygous mutant condition, oncogenes act in a dominant fashion: Only one copy of the activated gene need be present for transformation to occur.

With the viral oncogene in hand, researchers could create a probe for the gene and look within the DNA of

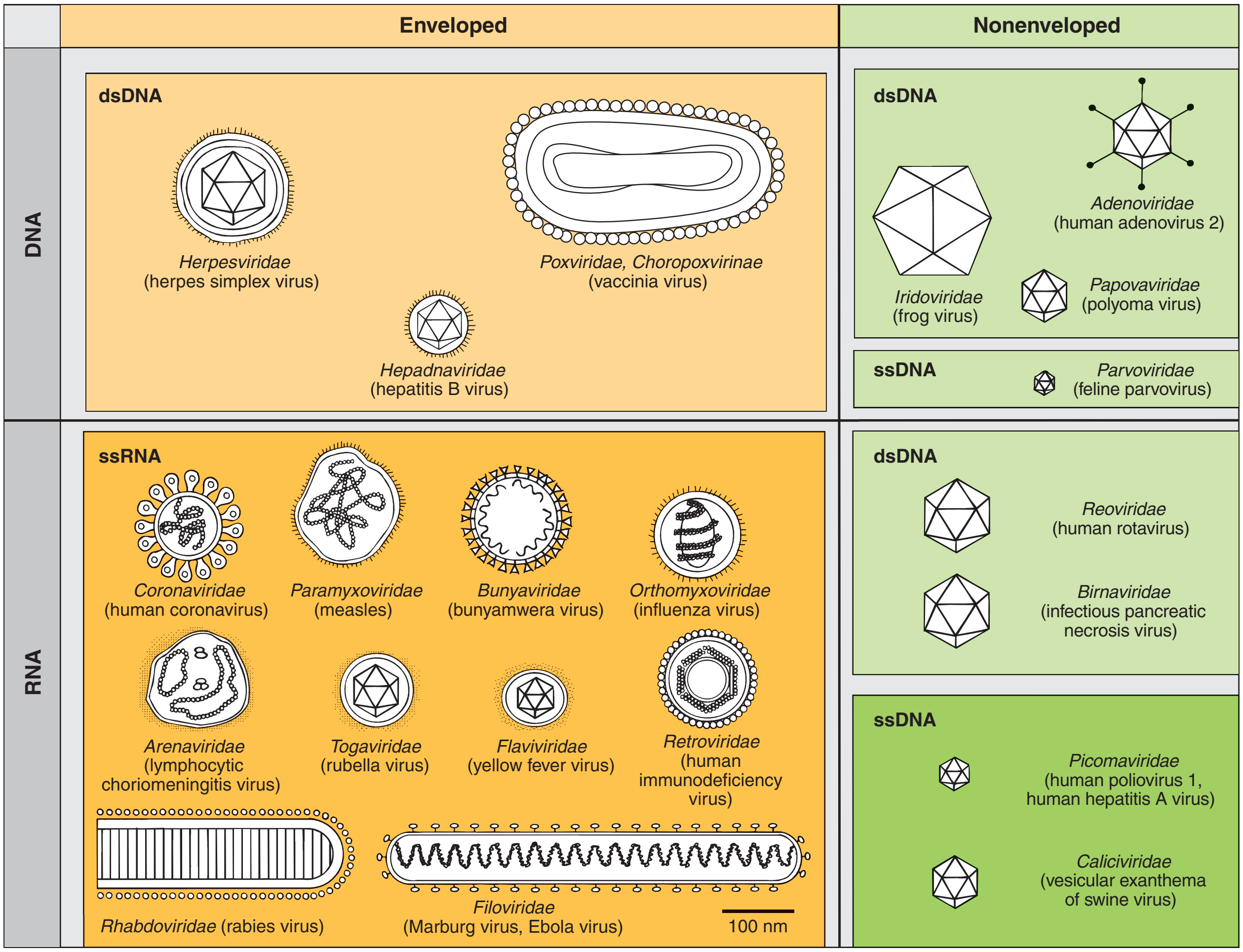

Enveloped
Nonenveloped
DNA
RNA
dsDNA
Herpesviridae
(herpes simplex virus)
Poxviridae, Choropoxvirinae
(vaccinia virus)
Hepadnaviridae
(hepatitis B virus)
dsDNA
Adenoviridae
(human adenovirus 2)
Iridoviridae
(frog virus)
Papovaviridae
(polyoma virus)
ssDNA
Parvoviridae
(feline parvovirus)
ssRNA
Coronaviridae
(human coronavirus)
Paramyxoviridae
(measles)
Bunyaviridae
(bunyamwera virus)
Orthomyxoviridae
(influenza virus)
Arenaviridae
(lymphocytic choriomeningitis virus)
Togaviridae
(rubella virus)
Flaviviridae
(yellow fever virus)
Retroviridae
(human immunodeficiency virus)
Rhabdoviridae (rabies virus)
Filoviridae
(Marburg virus, Ebola virus)
100 nm
dsDNA
Reoviridae
(human rotavirus)
Birnaviridae
(infectious pancreatic necrosis virus)
ssDNA
Picornaviridae
(human poliovirus 1, human hepatitis A virus)
Caliciviridae
(vesicular exanthema of swine virus)

**Figure 16.28** Representatives of families of animal viruses. The abbreviations *ss* and *ds* refer to single-stranded and double-stranded, respectively. (From R. I. B. Francki, et al., *Classification and Nomenclature of Viruses,* fifth report, 1991. Springer-Verlag, Vienna. Reprinted by permission.)

the host organism. To the surprise of virtually everyone, these oncogenes were found in untransformed cells. Since transforming viruses can function quite well as viruses without their oncogenes, and since cellular oncogenes have introns and viral oncogenes do not, geneticists generally accept the theory that these oncogenes originated in the host and were picked up, presumably as messenger RNAs, by the retroviruses. We believe that retroviruses pick up cellular genes by transcription readthrough, transcribing beyond the end of the integrated virus and producing a messenger RNA that is then incorporated into a viral particle after intron removal. Retroviruses can thus pick up genes adjacent to their point of integration.

To distinguish oncogenes within viruses and hosts, we prefix the name of a viral oncogene, such as *src,* with a v (v-*src*) and a cellular oncogene with a c (c-*src*). Cellular oncogenes within a nontransformed cell are called **proto-oncogenes.** How are proto-oncogenes induced to become oncogenes, and what do proto-oncogenes normally do in the cell?

### *Oncogene Induction*

Proto-oncogenes can be induced in at least three different ways. First, a mutation can cause a proto-oncogene to transform its host cell. For example, a *ras* proto-oncogene (see table 16.4) was converted to an oncogene when one codon, GGC (glycine), was converted to GTC (valine). Second, a proto-oncogene can be activated if it is moved to a region with a strong promoter or enhancer. Burkitt's lymphoma, for example, is associated with a translocation involving the proto-oncogene c-*myc,* which is normally located on chromosome 8. When translocated to

## Table 16.4 Some Oncogenes, Their Origins, and Their Protein Products

| Oncogene | Virus | Species of Origin | Gene Function |
|---|---|---|---|
| *abl* | Abelson murine leukemia virus | Mouse | Tyrosine kinase |
| *src* | Rous sarcoma virus | Chicken | Tyrosine kinase |
| *erbB* | Avian erythroblastosis virus | Chicken | Tyrosine kinase |
| *fms* | McDonough feline sarcoma virus | Cat | Growth factor |
| *mos* | Avian myeloblastosis virus | Chicken | Protein kinase |
| *sis* | Simian sarcoma virus | Woolly monkey | Growth factor |
| *Ha-ras* | Harvey murine sarcoma virus | Rat | GTP-binding protein |
| *Ki-ras* | Kirsten murine sarcoma virus | Rat | GTP-binding protein |
| *fos* | FBJ osteosarcoma virus | Mouse | Binds DNA |
| *myb* | Avian myeloblastosis virus | Chicken | Binds DNA |
| *erbA* | Avian erythroblastosis virus | Chicken | Binds DNA |
| *rel* | Reticuloendotheliosis virus | Turkey | Binds DNA |
| *jun* | Avian sarcoma virus 17 | Chicken | Binds DNA |

*Source:* Reprinted with permission from J. Marx, "What Do Oncogenes Do?," *Science,* 223:673–76, 1984. Copyright © 1984 American Association for the Advancement of Science.

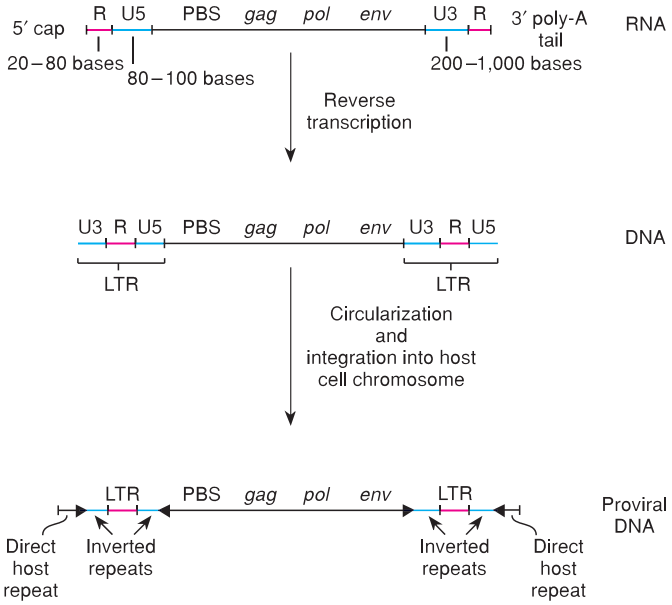

R
U5
PBS
*gag*
*pol*
*env*
U3
R
5′ cap
3′ poly-A tail
RNA
20–80 bases
80–100 bases
200–1,000 bases
Reverse transcription
U3
R
U5
PBS
*gag*
*pol*
*env*
U3
R
U5
DNA
LTR
LTR
Circularization and integration into host cell chromosome
LTR
PBS
*gag*
*pol*
*env*
LTR
Proviral DNA
Direct host repeat
Inverted repeats
Inverted repeats
Direct host repeat

**Figure 16.29** A retroviral RNA genome. R is a repeated sequence; U3 and U5 are unique sequences; PBS is the primer-binding site, LTR is the long terminal repeat; *gag, pol,* and *env* are viral genes. During the process of reverse transcription, the LTRs are created at the ends of the DNA. Direct host repeats are created when the viral DNA integrates into the host chromosome.

chromosome 14, c-*myc* is placed contiguous with the immunoglobulin IgM constant gene. This gene is very active in lymphocytes (as we will see later). Hence c-*myc* is now transcribed at a much higher rate than normal, resulting in cellular transformation. The c-*myc* gene normally occurs near a **fragile site,** a region of a chromosome that has a tendency to break. Many proto-oncogenes occur near fragile sites on chromosomes. The simple capture of a gene by a retrovirus might be enough for transformation, since the gene is brought under the influence of viral transcriptional control. However, not all genes captured this way are oncogenes. Third, a proto-oncogene can be activated if it is amplified. Several cases are known in which amplified genes (e.g., c-*ras* and c-*abl*) or genes on trisomic chromosomes are related to transformation.

Viral oncogenes can cause transformation by the same mechanisms. Either a mutation of the oncogene itself or the placement of the gene next to an active viral promoter can cause high levels of transcription of the oncogene and, hence, transformation of the cell. What are the gene products of these proto-oncogenes?

### *Oncogene Function*

We know that proto-oncogenes are important to the cell because they have been conserved evolutionarily. For example, c-*src* is found in fruit flies as well as in vertebrates; c-*ras* is found in yeasts and in human beings. They are all genes that can promote cell growth; they are specific transcription factors or other components of growth-stimulating signal pathways. The known protein products of oncogenes can be classified into at least four categories: tyrosine kinases, growth factors, GTP-binding proteins, and DNA-binding proteins (see table 16.4). As proto-oncogenes, they normally function at low levels; in transformed cells, they function at high levels. Can these proteins explain cancerous growth?

Tyrosine kinases are enzymes that add a phosphate group to tyrosine residues in proteins. Other kinases phosphorylate serine and threonine. (These three amino acids have OH groups available for phosphorylation.) Proteins that are phosphorylated at their tyrosine residues are involved in signal pathways, cytoskeleton shape (transformed cells are shaped differently than normal cells), and glycolysis (cancer cells tend toward the anaerobic glycolytic pathway). Overactivation of the cellular oncogene can result in inappropriate kinase activity, thereby changing many of the cellular activities and leading to cancer.

The c-*sis* oncogene encodes platelet-derived growth factor, which stimulates cells to grow. Its potential in transformation is obvious. GTP-binding proteins, the product of v-*ras,* for example, play a role in transmitting endocrine signals across membranes. Increased quantities of GTP-binding proteins can send continuous or amplified signals to certain cells and thus enhance growth. The v-*myc* gene product is a protein that binds to DNA; specific transcription factors can signal inappropriate transcription, also inducing transformation.

From the initial lesion in a gene to full-blown cancer normally takes many steps (clonal evolution; fig. 16.30). For example, in the colorectal cancer familial adenomatous polyposis (FAP), at least seven genetic changes are needed. Through these steps, a normal mass of cells passes through **hyperplasia,** an increased growth without any obvious change in cells; to **dysplasia,** in which overgrowth continues with changes in cell and nuclear structures (polyp formation); to the cancerous state, with invasion of surrounding tissues and metastasis. B. Vogelstein and colleagues have discovered many of the genetic changes involved in the formation of this cancer. First, the *APC* gene (*a*denomatous *p*olyposis *c*oli) mutates, leading through hyperplasia to dysplasia, a condition referred to as aberrant cryptic foci. Although the exact role of the APC protein is not known, it does bind β-catenin, which is involved in cell adhesion and activates the cyclin D1 promoter, exerting a direct effect on cellular proliferation. Thus, mutation of *APC* results in an accumulation of β-catenin, which then has effects on cell cycle progression and cell adhesion.

The next genetic change results in an early adenoma (a benign growth). Mutation of the *ras* oncogene leads to intermediate adenoma, due to the autonomous growth signals sent by the Ras GTP-binding protein. This is followed by late adenoma caused by the mutation of a gene in region q21 of chromosome 18, a gene called *deleted in colorectal cancer (DCC).* This gene codes for a transmembrane protein involved in the adhesion of cells to each other.

At this point, mutations resulting in the loss of *p53* protein result in full-blown cancer. Throughout this series of events, it is a cell from the previous state that mutates into the next state, consistent with our concept of clonal evolution. Although it may seem odd that so many mutations appear consecutively in the same cells, remember that mutations in some genes, such as *p53,* result in an overall higher mutation rate within cells. In one study, when cancer cells were compared with noncancerous progenitor cells, the cancer cells showed eleven thousand genetic changes.

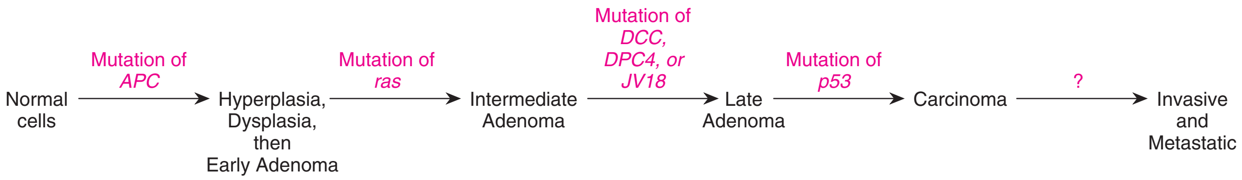


**Figure 16.30** Some of the known steps in converting a normal colon cell into a cancerous one. At least four known genes are involved, two oncogenes *(APC* and *ras)* and two tumor-suppressor genes *(DCC*—or *DPC4* or *JV18*—and *p53).* (Reprinted from *Cell*, Vol. 87, K. W. Kinzler and B. Vogelstein, "Lessons from Hereditary Colorectal Cancer," pp. 159–170, Copyright © 1996, with permission from Elsevier Science.)

In summary, all cancers share the following traits. First, they provide their own growth signals while ignoring inhibitory signals; in essence, cancer cells can grow without limit. Second, cancer cells avoid apoptosis. Third, tumor cells create new blood supplies by a process known as *angiogenesis;* new blood vessels grow in the tumors, allowing them unlimited increase in size. And finally, all malignant tumors have the capability of invasive growth and metastasis.

### Environmental Causes of Cancer

Environment plays a major role in carcinogenesis, and many environmental carcinogenic agents are known (table 16.5). Many of these agents are also mutagens (see chapter 12). Avoidable substances in the environment and the diet are estimated to cause 80 to 90% of all cancers, although the exact mechanisms by which these agents induce transformation are generally unknown. Perhaps the most effective cancer prevention strategy would be to avoid as many carcinogens from the environment as possible.

In the final section of the chapter, Immunogenetics, we look at another genetic system of transcriptional control. We try to answer the question: How does a single organism produce such a vast array of immunological protection?

**Table 16.5** **Carcinogenic Substances in the Environment**

| Carcinogen | Cancer Site(s) |
|---|---|
| Aromatic amines | Bladder |
| Arsenic | Liver, lung, skin |
| Asbestos | Lung |
| Benzine | Bone marrow |
| Chromium | Lung, nose, nasopharynx sinuses |
| Cigarettes | Lung |
| Coal products | Bladder, lung |
| Dusts | Lung |
| Ionizing radiation | Bone, bone marrow, lung |
| Iron oxide | Lung |
| Isopropyl oil | Nasopharynx sinuses, nose |
| Mustard gas | Lung |
| Nickel | Lung, nasopharynx sinuses, nose |
| Petroleum | Lung |
| Ultraviolet irradiation | Skin |
| Vinyl chloride | Liver |
| Wood and leather dust | Nasopharynx sinuses, nose |

## IMMUNOGENETICS

Vertebrates have evolved the ability to protect themselves against invading bacteria, viruses, and parasites and against their own cancer cells by creating an enormous amount of immune diversity with relatively few genes. Here we concentrate on the genetic control of **immunity,** the ability of an animal to resist infection. The foreign substance from the bacterium, virus, parasite, or cancer cell that evokes an immune response is called an **antigen.** The immune response itself is a complex interaction of various cell types, signaling pathways, and other components. The immune system of a mammal can destroy millions of different antigens without harming its own cells—quite an amazing accomplishment.

The two major components of the immune system are the B and T lymphocytes, white blood cells that originate in bone marrow and mature in either the bone marrow (B cells) or the thymus gland (T cells). The B cells are responsible for producing very specific proteins called **antibodies,** or **immunoglobulins (Igs),** which protect the organism from antigens in three general ways. Immunoglobulins can coat antigens so that they are more readily engulfed by phagocytes (white blood cells that engulf foreign material); immunoglobulins can combine with the antigens—for example, by covering the membrane-recognition sites of a virus—and thereby directly prevent their ability to function; or, in combination with *complement,* a blood component, immunoglobulins can cause the cell to die if the antigen is from an intact cell. B cells are the major component of **humoral immunity,** immunity controlled by antibodies in the serum and lymph; T cells are the major component of **cellular immunity,** immunity against infected cells.

Whereas the B cells produce immunoglobulins, one type of T cell is concerned with locating and destroying infected cells to prevent invading organisms from escaping detection within those infected cells. The **cytotoxic T lymphocytes** attack host cells infected by a virus, bacterium, or parasite. Thus, infected cells are destroyed before new viruses, bacteria, or parasites can be produced, helping to terminate the infection. Cytotoxic T lymphocytes recognize infected host cells by surface receptors called **T-cell receptors.** These receptors recognize an infected host cell by two aspects of the infected cell's surface: **major histocompatibility complex** (MHC) gene products, and antigens. All host cells have MHC components on their surfaces; an infected cell has the ability to cause part of the antigen to appear on its surface with the MHC protein, as if the MHC protein were "presenting" the antigen to the T-cell receptor (fig. 16.31).

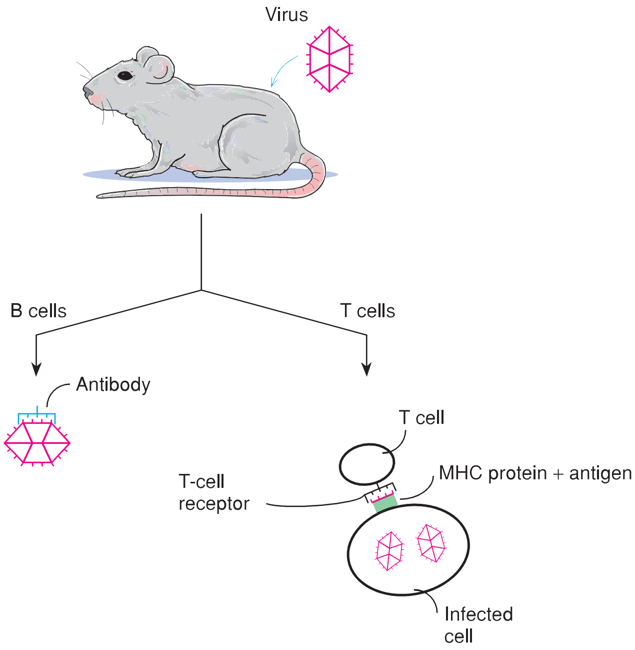


**Figure 16.31** When a mammal (e.g., mouse) is infected by a virus, part of the viral coat is recognized as an antigen, triggering an immune response. B cells produce antibodies that specifically attach to the viral antigen (humoral immunity). Infected cells "present" the antigenic part of the viral coat to the outside at the major histocompatibility complex protein on the cell surface. T-cell receptors recognize this MHC-antigen complex and then trigger the destruction of the infected cell (cellular immunity).

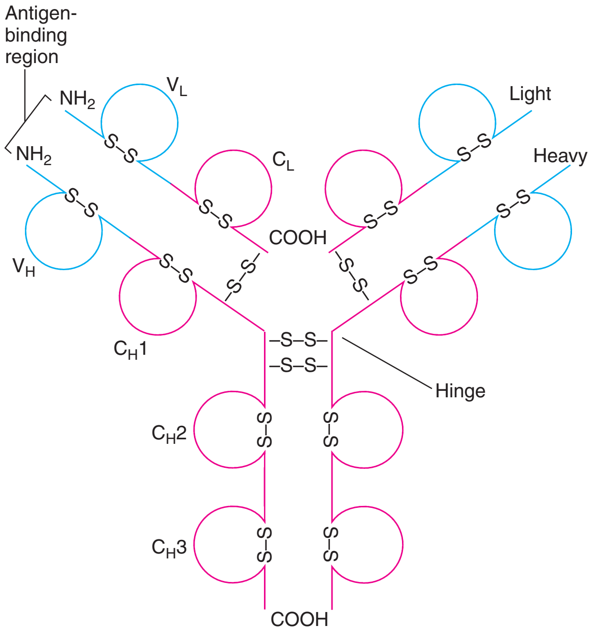


**Figure 16.32** Schematic view of an immunoglobulin protein (IgG). V = variable region; C = constant region; L = light chain; H = heavy chain. The S-S bonds are sulfhydryl bridges across two cysteines. The $NH_2$ ends of the molecule form the antigen-recognition parts. The internal sulfhydryl bonds roughly mark areas called domains, two each on the light chains and four each on the heavy chains. The heavy chain also has a hinge domain. Similar domains are found in the T-cell receptors and the MHC proteins. These domains indicate the evolutionary relatedness of these three types of molecules.

The dual attack by B and T cells has three main components of genetic interest: antibodies (immunoglobulins), T-cell receptors, and products of the major histocompatibility complex. These three protein families are evolutionarily related to each other, and each provides a diversity of protein products.

## *Immunoglobulins*

Immunoglobulins, produced by the B cells, are large protein molecules composed of two identical light polypeptide chains (about 214 amino acids) and two identical heavy chains (about 440 amino acids), held together by sulfhydryl bonds (fig. 16.32). Each polypeptide chain has a variable and a constant region of amino acid sequences. The variable regions recognize the antigens and thereby give specificity to the immunoglobulins (fig. 16.33). There are five major types of heavy chains (γ, α, μ, δ, and ε), giving rise to five types of immunoglobulins: IgG, IgA, IgM, IgD, and IgE. Each has slightly different properties; for example, only IgG can cross the placenta, giving immunity to the fetus. In addition, every immunoglobulin has one of two types of light chains, κ or λ (kappa or lambda).

Mutations of the constant region of the chains, called **allotypes,** follow the rules of Mendelian inheritance. In the variable region, however, called the **idiotypic variation,** diversity is much greater than two alleles per individual. The average individual has the potential to express between $10^6$ and $10^9$ different immunoglobulins, each with a different amino acid sequence. The lower limit, $10^6$, is arrived at through the study of persons with multiple myeloma, a malignancy in which one lymphatic cell divides over and over until it makes up a substantial portion of that person's lymphocytes. From these persons, we can isolate a relatively purified immunoglobulin that is the product of a single clone of cells and is referred to as a **monoclonal antibody.** A very low proportion of a normal person's lymphocytes produces any one specific immunoglobulin.

Multiple myeloma cells can be fused to spleen cells. The resulting cells, called **hybridomas,** which produce monoclonal antibodies, can be perpetuated in tissue culture indefinitely, thus providing a ready supply of specific monoclonal antibodies. Recent work with hybridomas has allowed us to locate, isolate, and sequence immunoglobulin genes. How can one genome produce $10^9$ different antibody molecules?

## Antibody Diversity

Since the mammalian genome does not have $10^9$ genes ($10^5$ genes is a better estimate), different models were suggested to explain antibody diversity. In 1965, W. J. Dreyer and J. C. Bennett suggested that a given chain of an immunoglobulin was not the result of one gene, but of a combination of genes, one for the constant region and one for the variable region. In addition, they suggested that a particular organism, in its haploid genome, had only one constant gene but several hundred or thousand variable genes. The final product would be the result of the action of a combination of one of the variable genes and the constant gene. Modern recombinant DNA technology has verified the essence of that model.

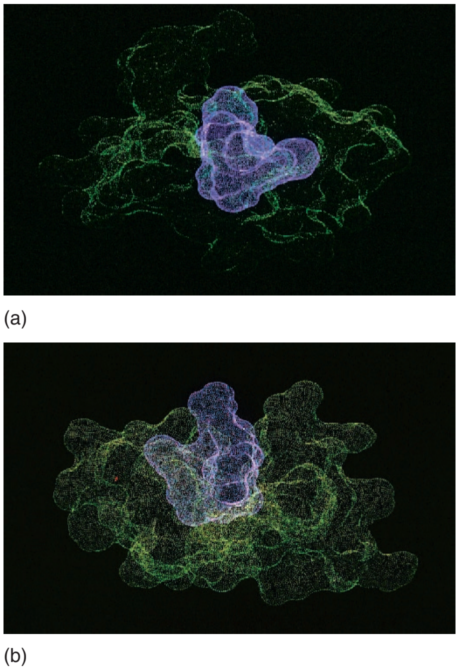


**Figure 16.33** A computer-generated view of the interaction of an antigen-binding region of an immunoglobulin *(green)* with an antigen *(purple:* in this case, the hormone angiotensin II, composed of only eight amino acid residues). Note how the antigen fits into the variable end of the immunoglobulin, as an apple fits into a cupped hand. (*a*) Top view; (*b*) side view.
(Courtesy of L. Mario Amzel.)

In reality, several genes contribute to form the variable regions of the heavy and light chains, given that we are using the word *gene* for DNA segments that code for a part of the final heavy or light chain of the immunoglobulin. Genes for the κ, λ, and heavy chains are located on chromosomes 2, 22, and 14, respectively, in human beings. Each is a multigene complex. Let us examine the κ light-chain gene complex as an example of how the DNA must be modified to produce the final protein product (fig. 16.34).

The first step in DNA rearrangement is the joining of a V (variable) and a J (joining) gene in a B cell (fig. 16.34), a process called V-J joining. Since any one of eighty V genes can combine with any one of five J genes, four hundred different combinations are possible (80 × 5). Since we expect this to be another example of site-specific recombination, as we saw with phage λ integration in chapter 14, recombinational signal sequences must be flanking all genes so that any two can be moved next to each other. Through DNA sequencing, these signals, termed *recombination signal sequences,* have been determined to be a heptamer (seven bases) and a nonamer (nine bases), separated by twelve bases on one side and twenty-three bases on the other (known as the *12-23 rule;* fig. 16.35).

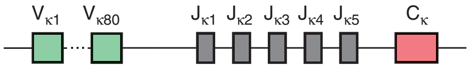


**Figure 16.34** The complex for the human κ light chain is composed of about eighty variable genes ($V_{\kappa1}$–$V_{\kappa80}$), five joining genes ($J_{\kappa1}$–$J_{\kappa5}$), and one constant gene, $C_{\kappa}$, in the undifferentiated cell (germ line). The final κ light chain will be composed of the products of one variable gene, one joining gene, and the constant gene.

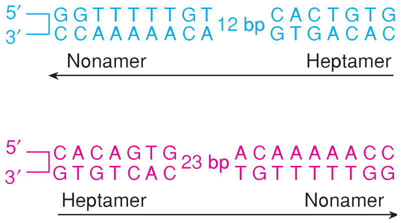


**Figure 16.35** In order for V-J joining to take place (site-specific recombination), there must be signals at the V side and at the J side. One signal is a heptamer (seven base pairs) and a nonamer (nine base pairs) separated by twenty-three base pairs, and the other signal is the same heptamer and nonamer separated by twelve base pairs, in reverse orientation.

Two proteins, RAG1 and RAG2 (from *r*ecombination *a*ctivating *g*enes 1 and 2), form a recombinase enzyme capable of recognizing 12–23 signals and producing double-strand breaks in DNA at the junction of coding (V and J) regions and signal sequences (12–23 regions; fig. 16.36). Recognition of the 12–23 signals is done with the help of HMG1 or HMG2 (*h*igh *m*obility *g*roup proteins 1 and 2), proteins that bind DNA and bend it, enhancing the activity of the RAG recombinase. The recombinase brings together one of the variable regions and one of the joining regions and links them in a process that also generates a circle of DNA with the intervening segments, which is then eliminated (fig. 16.36).

The RAG recombinase nicks each DNA strand at its recombination signal sequence. Then by a transesterification (shifting a phosphodiester bond), the recombinase forms a hairpin and a blunt end, a double-strand break of the DNA (fig. 16.37). The remainder of the DNA processing is done with double-strand break repair enzymes that repair radiation-induced DNA damage (fig. 16.37; see chapter 12). The result is an eliminated circle of intervening genes and a chromosome with a V region adjacent to a new J region.

The point of crossover at the V-J junction is itself variable, generating **junctional diversity.** Not only are any two V and J genes capable of coming together, but also the sequence at the junction of the two genes can vary. For example, we see in figure 16.38 that the junction in the protein at amino acids 95 and 96 can be Pro-Trp, Pro-Arg, or Pro-Pro, depending on exactly where the crossover occurred.

In figure 16.39, we see the DNA after $V_{\kappa 50}$ and $J_{\kappa 4}$ join. This gene is now transcribed. The region between $J_{\kappa 4}$ and $C_{\kappa}$ (the constant gene) is then removed by RNA splicing, leaving the final messenger RNA product, which is then translated into a κ light chain. In this cell, the homologous κ region is repressed as well as both λ regions, a phenomenon known as **allelic exclusion.** Thus, this cell produces only one light chain, the $V_{\kappa 50}$-$J_{\kappa 4}$-$C_{\kappa}$ protein.

Similar types of events take place in the heavy-chain gene and the λ light-chain gene if it has been activated. There are some differences, however (fig. 16.40). The λ complex in human beings has only two variable genes, with four J genes and one C gene. The heavy-chain

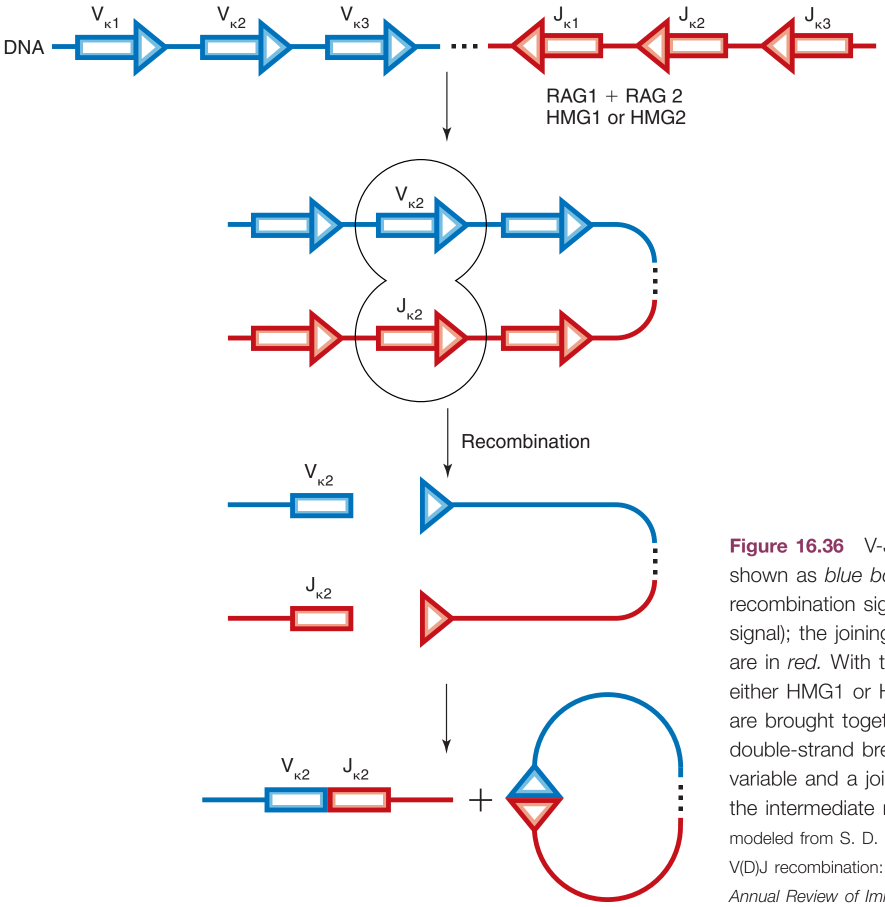


**Figure 16.36** V-J joining. The variable genes are shown as *blue boxes* with *blue arrowheads* as the recombination signal sequences (the 12 spacer signal); the joining genes with their 23 spacer signals are in *red.* With the RAG1-RAG2 recombinase and either HMG1 or HMG2, a variable and a joining gene are brought together. Recombination results in double-strand breaks that are then repaired so that a variable and a joining gene are spliced together and the intermediate material is released. (Source: Diagram modeled from S. D. Fugmann, et al., "The RAG proteins and V(D)J recombination: Complexes, ends, and transposition," *Annual Review of Immunology,* 18:495–527, 2000.)

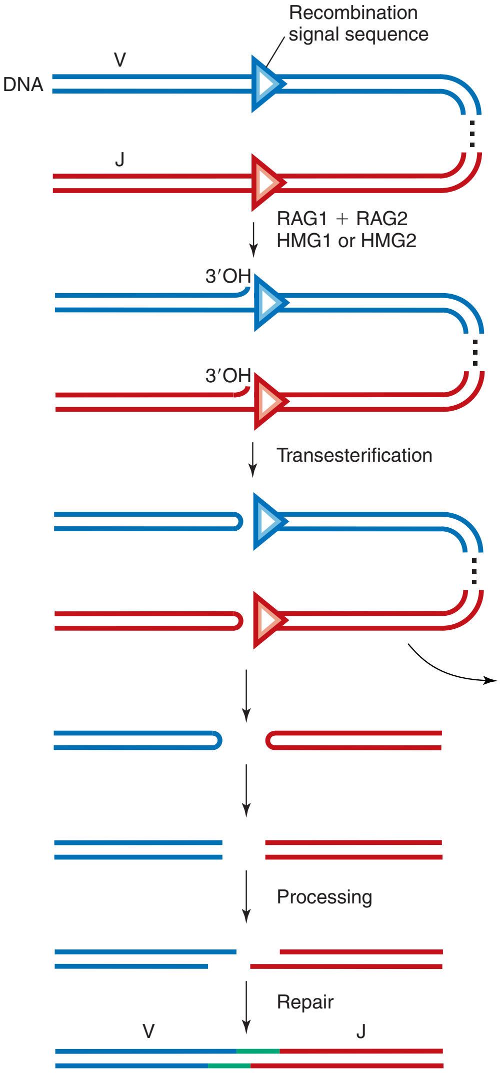


**Figure 16.37** The mechanism of site-specific recombination between a variable and joining gene. The RAG recombinase, with the help of HMG1 or HMG2, recognizes the recombination signal sequences *(red* and *blue arrowheads)*. A nick at each signal is made, producing 3′OH ends; transesterification then forms hairpin loops and recombinational signals with double-strand breaks. The hairpin loops are brought together, opened, and repaired, with some processing taking place. This creates junctional diversity, including crossover point variability and N segments (to be discussed later). The enzymes responsible for the processing are repair enzymes.

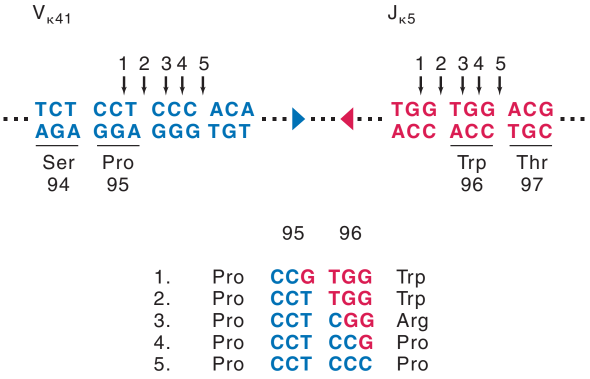


**Figure 16.38** Variability in crossing over during V-J joining generates junctional diversity. In this case, $V_{\kappa 41}$ and $J_{\kappa 5}$ are shown. Amino acid codons 94 and 97 are always the same, TCT and ACG, respectively, as are the first two bases of codon 95, CC. Depending on the exact point of crossover, five different codon pairs can be generated. Codons for proline (Pro) are always the first in each pair (95), but codons for tryptophan (Trp), arginine (Arg), or Pro are all possible second codons (96). Matching numbered arrows indicate crossover points for the five possibilities. (Source: Data from E. E. Max, et al., "Sequences of five potential recombination sites encoded close to an immunoglobulin k constant region gene," *Proceedings of the National Academy of Sciences,* 76:3450–54, 1979.)

complex has about one hundred to three hundred V genes, nine J genes, and the five C genes of the five major types (γ, α, μ, δ, ε). In addition, heavy-chain regions have another set of genes, called diversity (D) genes. At least five such genes are in the human heavy-chain complex, and they add still another variable region to the final protein. In the heavy chain, D-J joining first takes place, then V-DJ joining; lastly, splicing creates the final heavy-chain product (fig. 16.41).

As we pointed out earlier, the final form of the heavy-chain protein in human beings has five regions or domains—$C_H3$, $C_H2$, hinge, $C_H1$, and variable region (see fig. 16.32). Each of the constant regions, as well as the hinge region, comes from its own exon (fig. 16.42). (The variable region, of course, comes from the extensive recombination just described: fig. 16.41.) The heavy chain is thus another example of the relationship between exon structure and domain function, a topic we discussed in chapter 10. Heavy-chain structure would support the exon shuffling view (*introns early*).

V-J, D-J, and V-DJ joining, collectively called **V(D)J joining,** are the only known examples of site-specific recombination in vertebrates. The genes responsible are active only in pre-B and pre-T cells.

In addition to V(D)J joining, junctional diversity is also added during heavy-chain recombination by the addition of nucleotides in a template-free fashion. In other words, added nucleotides, called **N segments,** appear at the joining junctions; they are not specified in the DNA. For example, in one case, the sequence GTGGGGGCC (three codons long) was found at a D-J junction, but not seen in the undifferentiated (germ-line) genome. Mice that lack the gene for terminal deoxynucleotide transferase lack these N segments, implicating that gene in the process of N-segment formation. The enzyme adds nucleotides at the 3′ ends of the DNA strands; these free ends are created during V(D)J joining. The enzyme is found in high levels in immature lymphocytes.

There is a final way in which variability is generated. Sequencing studies indicate that mutation occurs in variable regions after recombination has taken place. The mechanism of this specific mutagenesis, called **somatic hypermutation,** is not known. Given the number of variable, constant, joining, and diversity genes, as well as the variation at the joining junctions, it is easy to see how $10^9$ different immunoglobulin combinations could be generated (table 16.6).

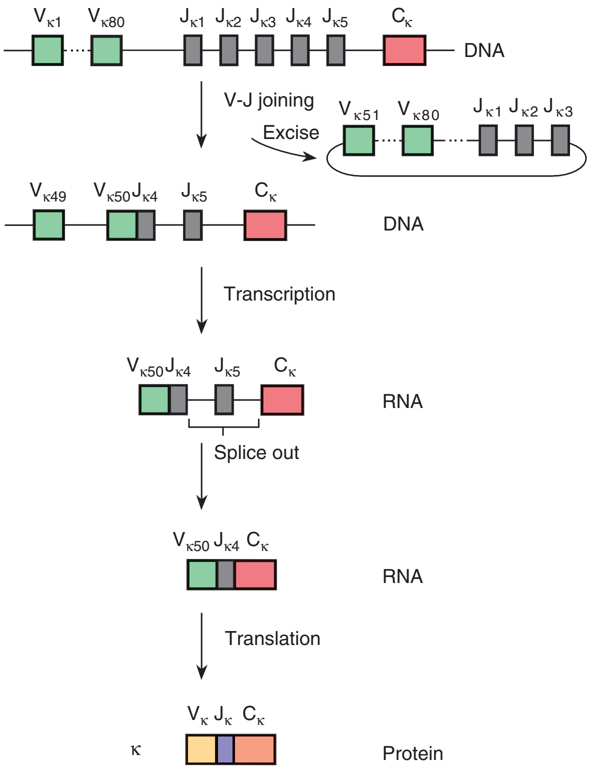

Vκ1
Vκ80
Jκ1 Jκ2 Jκ3 Jκ4 Jκ5
Cκ
DNA
V-J joining
Excise
Vκ51 Vκ80
Jκ1 Jκ2 Jκ3
Vκ49
Vκ50Jκ4
Jκ5
Cκ
DNA
Transcription
Vκ50Jκ4 Jκ5 Cκ
RNA
Splice out
Vκ50 Jκ4 Cκ
RNA
Translation
Vκ Jκ Cκ
κ
Protein

**Figure 16.39** V-J joining in the human κ region. In this example, $V_{\kappa 50}$ is joined to $J_{\kappa 4}$ and then to $C_\kappa$. First, V-J joining takes place using the heptamer-nonamer signals shown in figure 16.35. Then the region from the $V_{\kappa 50}$ to the $C_\kappa$ genes is transcribed. Splicing the RNA removes the region containing the extra J gene, $J_{\kappa 5}$. The final RNA, containing $V_{\kappa 50}$-$J_{\kappa 4}$-$C_\kappa$, is then translated into the κ light chain.

**Table 16.6** **Three Hundred Immunoglobulin Genes Can Generate 1.8 Billion Different Antibodies**

| Source | Factor |
|---|---|
| Light Chains | |
| V genes | 40× |
| J genes | 5× |
| V-J recombination* | 10× = 2,000× |
| Heavy Chains | |
| V genes | 200× |
| D genes | 5× |
| J genes | 9× |
| V-D, D-J recombination* | 100× = 900,000× |
| Total | 2,000 × 900,000 = 1.8 billion |

*Source:* Data modified from P. Leder, "The genetics of antibody diversity," *Scientific American,* 102–15, May 1982.

* Junctional diversity, N-segment formation, and hypermutability.

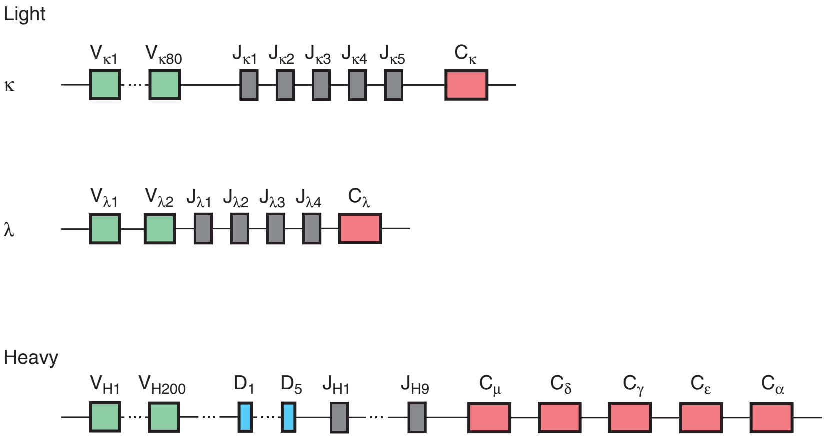

Light
κ
Vκ1
Vκ80
Jκ1 Jκ2 Jκ3 Jκ4 Jκ5
Cκ
λ
Vλ1 Vλ2 Jλ1 Jλ2 Jλ3 Jλ4 Cλ
Heavy
VH1 VH200
D1 D5 JH1
JH9
Cμ Cδ Cγ Cε Cα

**Figure 16.40** Arrangement of the genes in the light and heavy chains of the human immunoglobulin complexes.

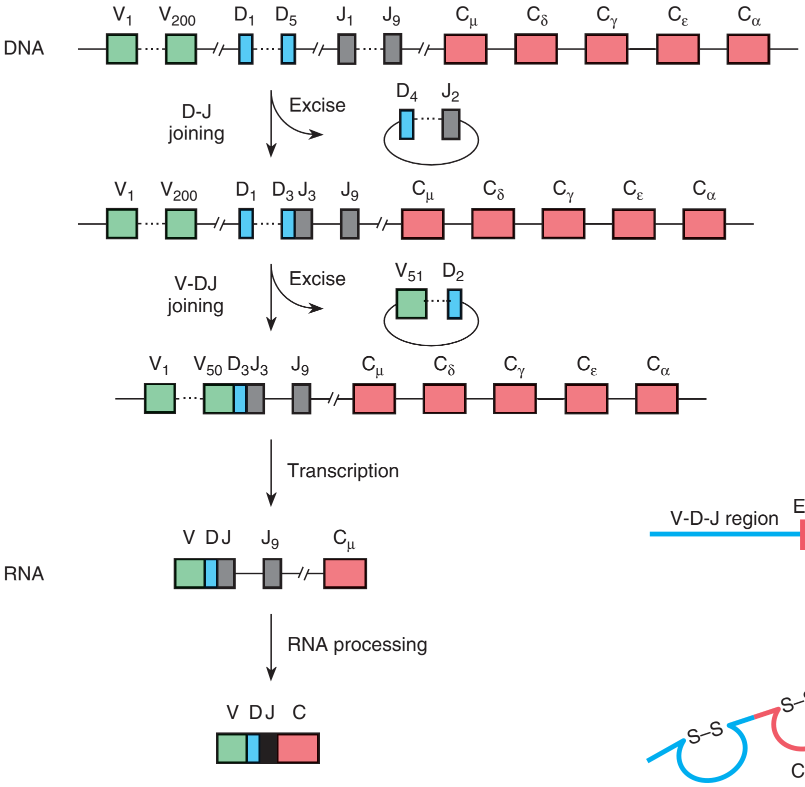


**Figure 16.41** Formation of an immunoglobulin heavy chain (IgM). First, D-J joining takes place, followed by V-DJ joining. In each case, intervening DNA is spliced out by site-specific recombination. Then, as in light-chain formation (fig. 16.39), the modified region is transcribed; RNA processing (splicing) then brings the final regions together, which, when translated, form the V-D-J-C heavy chain. (Source: Data from F. W. Alt, et al., "Development of the primary antibody repertoire," *Science,* 238:1079–87, 1987.)

## *T-Cell Receptors and MHC Proteins*

As we mentioned earlier, genetic diversity also exists in the T-cell receptors and the major histocompatibility complex (MHC). From its function (recognizing both the antigen and the MHC "self" gene product), it seems evident that the T-cell receptor must show the same type of diversity that immunoglobulins have. In fact, the T-cell receptor genes are very similar to the immunoglobulin genes. T-cell receptors are composed of $\alpha$ and $\beta$ subunits; there are V, J, and C components of the $\alpha$ subunit and V, J, D, and C components of the $\beta$ subunits (fig. 16.43). In a sampling of T-cell receptors from one individual, approximately one million different $\beta$ chains and 25 different $\alpha$ chains were found, yielding approximately 25 million (1 million × 25) different T-cell receptors.

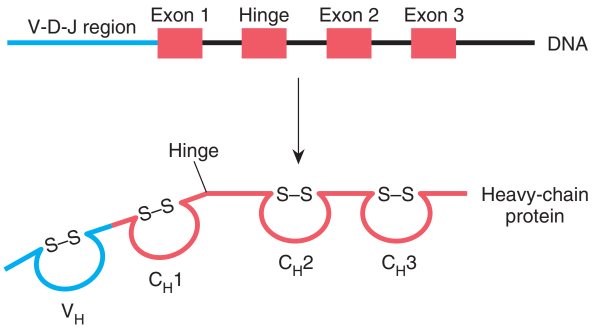


**Figure 16.42** The constant portion of heavy-chain genes is made of four domains, each transcribed from its own exon.

The major histocompatibility complex (MHC) region (also known as the human leukocyte antigen or HLA region in people) comprises a region of 3.6 million base pairs with 224 identified genetic loci. The genes are generally referred to as class I, II, and III genes. Class III genes code for proteins in the complement system, which is involved in the destruction of foreign cells. Class I and II genes code, in part, for proteins that present antigens to T cells. That is, class I and II proteins form structures with grooves on their surfaces that are shaped to hold small polypeptides. These polypeptides can be normal breakdown products of cellular metabolism in healthy cells ("self" proteins) or parts of foreign invaders or their gene products in infected cells. Although similar, the two types of MHC proteins are found in different places and serve somewhat different functions.

Class I MHC proteins consist of a membrane-bound $\alpha$ chain and a second chain called $\beta_2$ macroglobulin (fig. 16.44*a*). Class II MHC proteins consist of an $\alpha$ and $\beta$ chain (fig. 16.44*b*). These two proteins present antigens somewhat differently. In the class I molecules, the

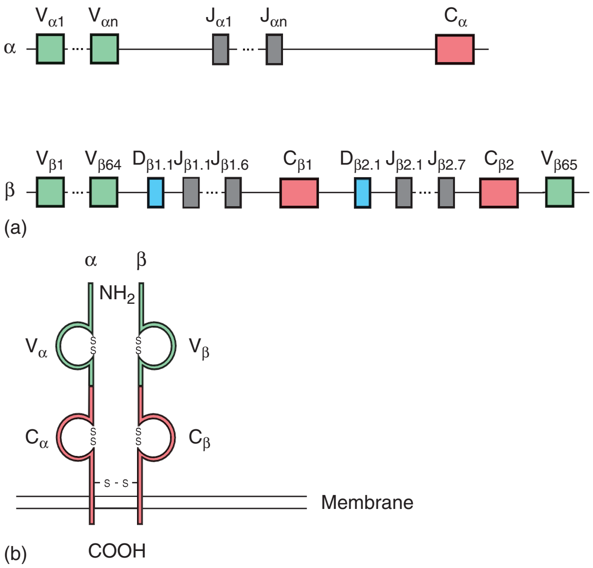

$V_{\alpha 1}$
$V_{\alpha n}$
$J_{\alpha 1}$
$J_{\alpha n}$
$C_{\alpha}$
α
$V_{\beta 1}$
$V_{\beta 64}$
$D_{\beta 1.1}$
$J_{\beta 1.1}$
$J_{\beta 1.6}$
$C_{\beta 1}$
$D_{\beta 2.1}$
$J_{\beta 2.1}$
$J_{\beta 2.7}$
$C_{\beta 2}$
$V_{\beta 65}$
β
(a)
α
β
$NH_2$
$V_{\alpha}$
$V_{\beta}$
$C_{\alpha}$
$C_{\beta}$
S - S
Membrane
(b)
COOH

**Figure 16.43** The T-cell receptor is made up of two protein chains, α and β, anchored in the cell membrane. Each protein is similar to an immunoglobulin chain and is created the same way, with V-J types of joining taking place. (*a*) The α gene complex is composed of numerous V and J genes and one constant gene. The β gene complex has V, D, J, and C genes. (*b*) Each protein chain has two domains, similar to the domains of the immunoglobulins, indicating a common evolutionary ancestry.

groove is bound on both sides so that the presented polypeptide is small and defined (fig. 16.45). In the class II protein, the groove is unbound, allowing for a longer polypeptide (fig. 16.46).

Class I MHC proteins are found on almost all cells. The polypeptides that an infected cell presents with the MHC I proteins come from the breakdown of proteins within the cytoplasm of the cell. Peptides are targeted for breakdown when a ubiquitin molecule binds to the protein, sending a cellular signal that the protein is to be degraded. (Ubiquitin is a small polypeptide of 76 amino acid residues, highly conserved in eukaryotes.) The ubiquitin-tagged protein is unfolded, in an ATP-dependent process, and then fed into a **proteasome,** a barrel-shaped cellular organelle for protein breakdown (fig. 16.47). Then the peptide fragments associate with two proteins, together called TAP (*t*ransporter for *a*ntigen *p*rocessing), that prevent further degradation of the peptide as well as transport the peptide into the endoplasmic reticulum, where the peptide binds to the class I MHC proteins. The MHC I proteins with antigen are then transported to the cell surface. Passing T cells, called killer T cells or CD8 T cells because of their CD8 receptor protein, recognize the foreign antigen presented by the MHC I protein and release substances that kill the infected cells (fig. 16.48*a*). (White blood cells are classified by their surface antigens; the CD designation comes from *c*luster of *d*ifferentiation antigens used for this purpose.)

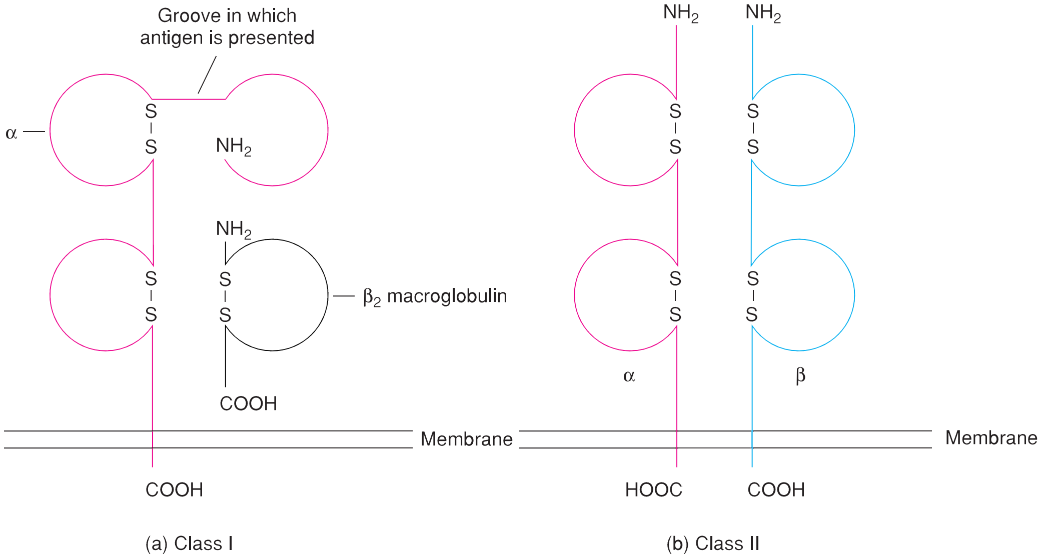

Groove in which antigen is presented
α
S
S
$NH_2$
$NH_2$
S
S
S
S
$\beta_2$ macroglobulin
COOH
Membrane
COOH
(a) Class I
$NH_2$
$NH_2$
S
S
S
S
S
S
S
S
α
β
Membrane
HOOC
COOH
(b) Class II

**Figure 16.44** The major histocompatibility complex (MHC) class I protein (*a*) is composed of two protein chains. The α chain is composed of three domains similar to the immunoglobulins and T-cell receptors. The second chain is $\beta_2$ macroglobulin. The MHC class II protein (*b*) is composed of an α and a β chain. The MHC proteins present antigens to the T-cell receptors to signal that a foreign agent has invaded the cell.

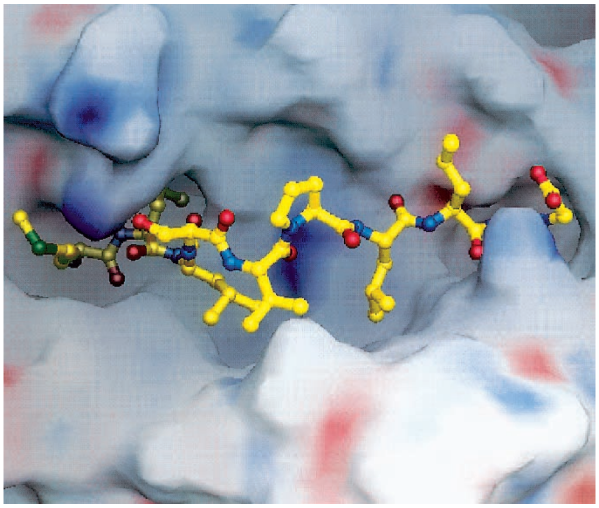


**Figure 16.45** A three-dimensional computer model of the antigen-presenting site of an MHC class I protein. The presented peptide is nine amino acid residues long and internally bound at each end. (Courtesy of Don C. Wiley.)

Class II MHC proteins are found only on cells involved in the immune system, such as macrophages and B lymphocytes. These cells are most likely to have encountered foreign objects like bacteria or parasites by having engulfed them. The MHC II proteins present foreign antigens not from the cytoplasm, but from *endosomic vesicles* within the cells. These vesicles form by budding from the cell surface and often contain foreign proteins and protease enzymes. The MHC II proteins migrate into the vesicles, where they pick up foreign polypeptides and then migrate to the cell surface. The response of passing T cells to the presentation by MHC II proteins is different from the response to MHC I proteins. The MHC II proteins with antigens are recognized by helper T cells, also called CD4 cells because of their surface receptor protein. Rather than kill cells such as infected macrophages that are useful in the immune system, the helper T cells stimulate the macrophages to destroy the foreign bacteria in their endosomic vesicles. The helper T cells also activate antibody-producing B cells. CD4 cells are the prime targets of the HIV virus, making individuals with AIDS very prone to bacterial infections and other immune problems (box 16.3).

One last point is worth mentioning about the MHC system. The loci for MHC proteins do not have V(D)J joining to produce the high levels of variability found in B and T cells. The MHC loci are, however, very variable, with many alleles. (That variability is one reason why organ transplants are usually rejected without immunosuppressive therapy.) Each individual can have only two alleles at each locus, but hundreds of alleles exist in any population. Presumably, the different alleles allow for somewhat different affinities for different antibodies. They may have been selected over evolutionary time to give certain individuals in a population more chances to be able to identify and eliminate foreign substances.

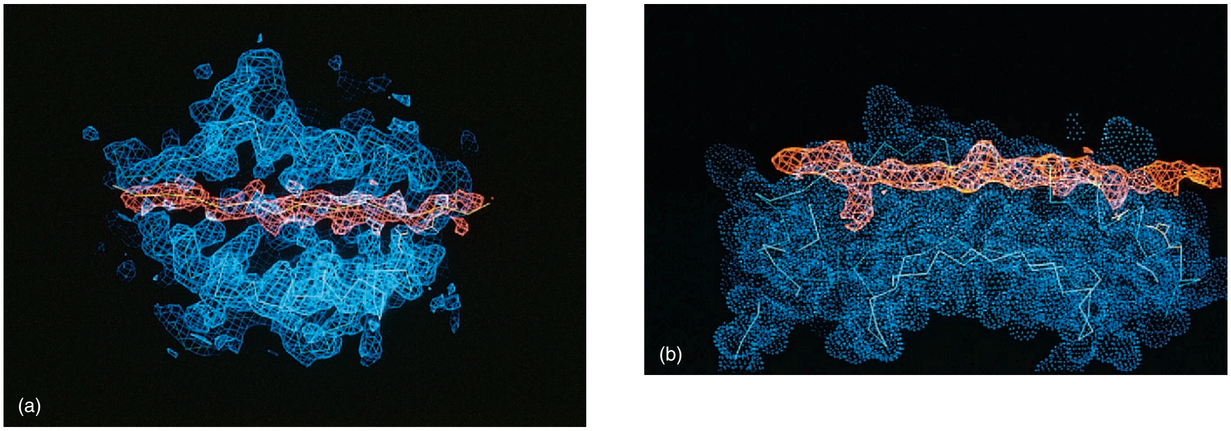


**Figure 16.46** A three-dimensional computer model of the antigen-presenting site of an MHC class II protein. The presented peptide is fifteen amino acid residues long and is not internally bound at each end. In (*a*), the view as seen by the T cell; (*b*) is a side view. The presented peptide is shown in *red;* the electron surface of the MHC protein is *blue.* (From: J. H. Brown, et al., *Nature* 364:33–39, 1993, fig. 4, p.35.)

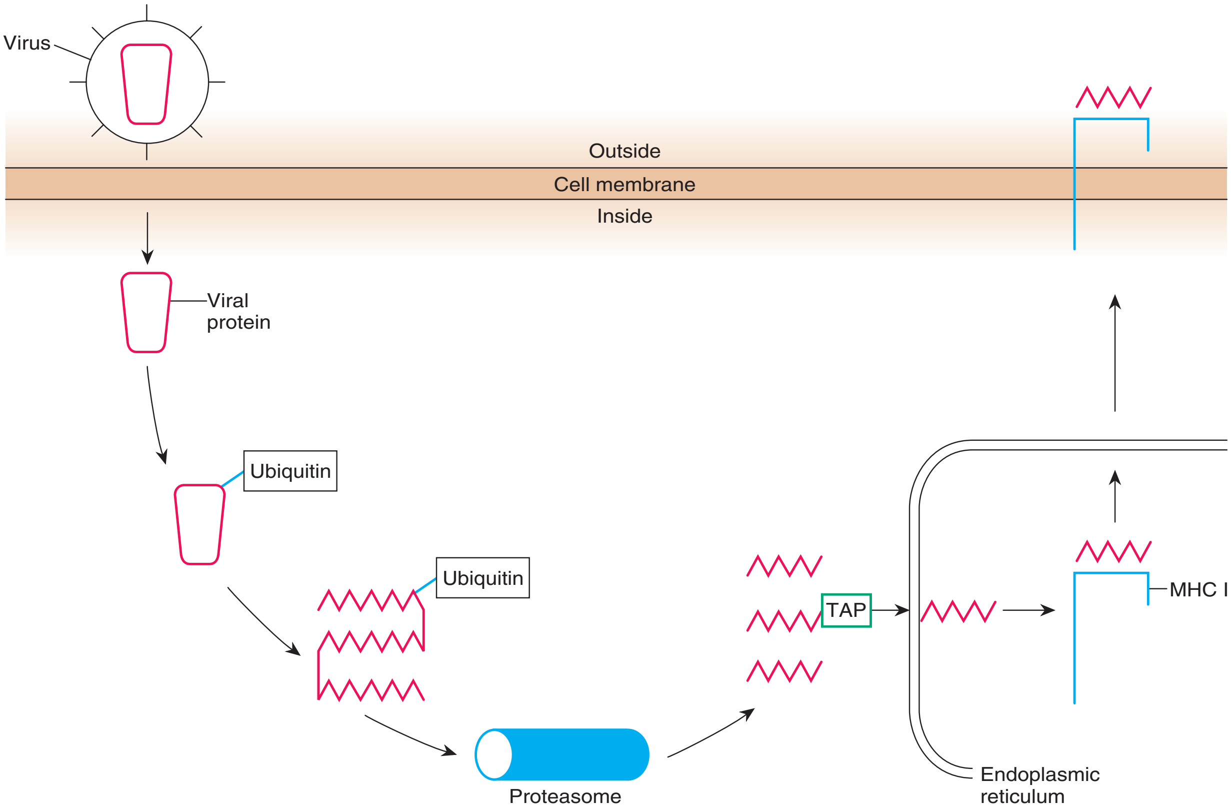

Virus
Outside
Cell membrane
Inside
Viral protein
Ubiquitin
Ubiquitin
Proteasome
TAP
MHC I
Endoplasmic reticulum

**Figure 16.47** How the MHC I protein obtains foreign peptide to display at the cell surface. In this example, a virus attacks a cell. The viral protein is recognized as foreign and is tagged with ubiquitin. The tagged protein is then unfolded and fed into a proteasome. With the aid of TAP, a piece of the degraded protein enters the endoplasmic reticulum, where it combines with the MHC I protein, which is then transported to the cell surface.

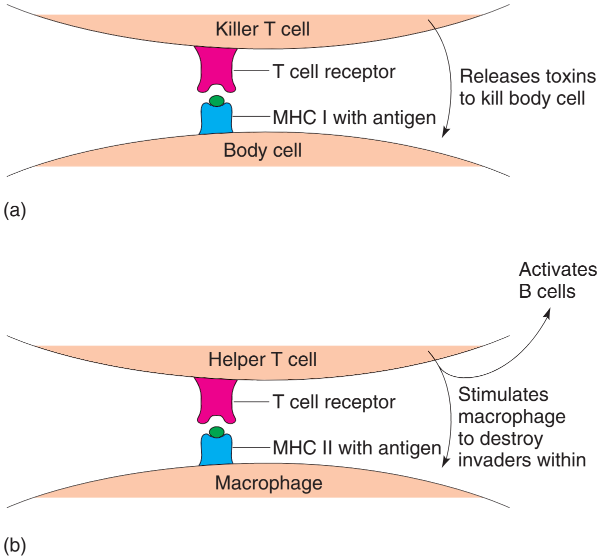

Killer T cell
T cell receptor
MHC I with antigen
Body cell
Releases toxins to kill body cell
(a)
Activates B cells
Helper T cell
T cell receptor
MHC II with antigen
Macrophage
Stimulates macrophage to destroy invaders within
(b)

**Figure 16.48** MHC class I and II proteins are found on different types of cells and recognized by different types of T cells. In (*a*), a normal body cell presents a foreign antigen in an MHC class I protein that a passing killer T cell recognizes. The T cell then releases toxins that kill the infected cell. In (*b*), a macrophage presents an antigen in an MHC class II protein that a helper T cell recognizes. The T cell stimulates the macrophage to destroy its invaders and also stimulates a B-cell reaction.

## BOX 16.3

## Biomedical Applications

### *AIDS and Retroviruses*

In 1983, Robert C. Gallo of the National Cancer Institute and Luc Montagnier of the Pasteur Institute of Paris co-discovered HIV—the human immunodeficiency virus, causative agent of acquired immune deficiency syndrome, or AIDS (fig. 1). HIV is a retrovirus causing a disease first diagnosed in 1981 among young male homosexuals in the United States.

The AIDS virus attacks helper T cells; a particular protein on the surface of these T cells, called CD4, is a receptor for the HIV virus coat protein, gp120 (fig. 2). A secondary receptor, the protein CCR5, is also needed for the virus to gain entry into the cell. (CCR5 refers to cysteine-cysteine linked *c*ytokine *r*eceptor 5.) HIV also attacks macrophages. With destruction of the T cells, a person's immune system loses the ability to fight off common diseases. Persons who develop the disease frequently fall victim to opportunistic diseases such as pneumonia caused by the protozoan *Pneumocystis carinii;* Kaposi's sarcoma, a rare cancer found in people taking immunosuppressive drugs; and several other conditions, normally rare except in people with suppressed immune systems. These conditions collectively became known as the acquired immune deficiency syndrome.

**EPIDEMIOLOGY**

AIDS has spread throughout the world. A 1959 blood sample from central Africa contained the first known human infection. By sequencing similar viruses in primates (simian immune deficient viruses, SIVs), researchers discovered that the common form of AIDS, caused by HIV-1, jumped from chimpanzees to human beings in the region of Gabon in western Africa. HIV-2, causing the less common form of AIDS, came from sooty mangabeys; SIVs have jumped to human beings at least seven times.

There seem to be two worldwide patterns in the spread of AIDS, which is not contracted by casual contact. In the New World, Australia, and Western Europe, homosexual men and intravenous drug users primarily spread the disease and are the groups at highest risk. In Africa and the Caribbean, the disease is spread primarily through heterosexual sex. Parts of southern Africa have infection rates between 16 and 32%; Eastern Europe, Asia, and North Africa have relatively low infection rates. In the United States, over 750,000 persons have the AIDS virus, with 350,000 deaths reported. Worldwide,

Robert C. Gallo (1937– ).
(Courtesy of Dr. Robert Gallo.)

Luc Montagnier (1932– ).
(Courtesy of Dr. Luc Montagnier.)

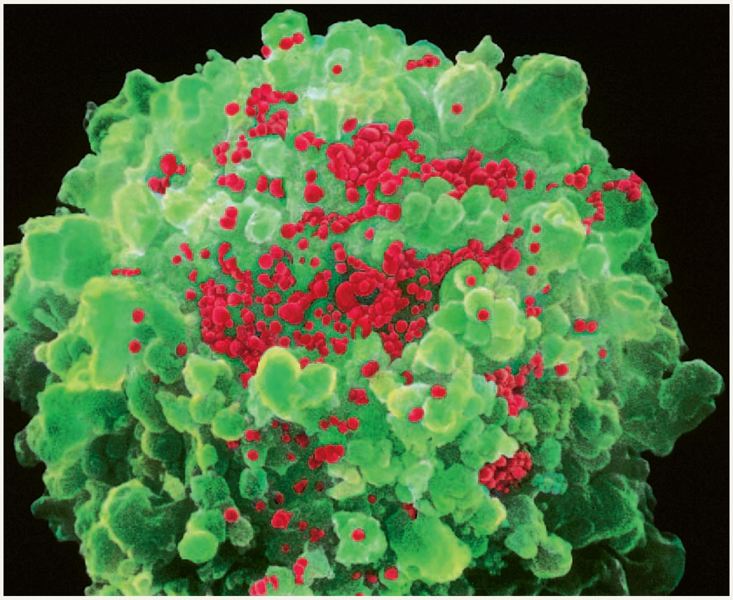


**Figure 1** Scanning electron micrograph of a T-lymphocyte *(green)* infected with the AIDS virus. Small spherical structures *(red)* on the surface of the cell are new virus particles budding off. (© NIBSC, Science Source/Photo Researchers, Inc.)

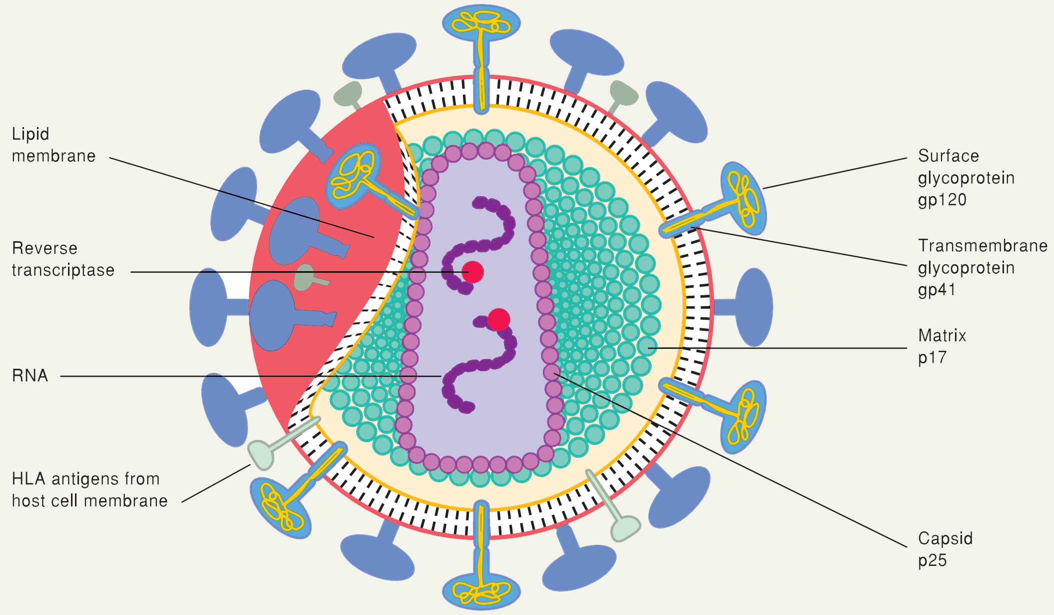

Lipid membrane
Reverse transcriptase
RNA
HLA antigens from host cell membrane
Surface glycoprotein gp120
Transmembrane glycoprotein gp41
Matrix p17
Capsid p25

**Figure 2** AIDS virion structure. Numbers associated with proteins are kilodalton masses (e.g., gp120 is a 120-kilodalton protein). (From Nester et al., *Microbiology: A Human Perspective,* 3rd edition. Copyright © 2001 The McGraw-Hill Companies, Inc. Reprinted with permission.)

over 33 million people are affected. Most of those who got the disease before 1990 have died. However, the infection rate seems to have peaked in the United States in 1985; the only area in which infections are increasing is through heterosexual sex.

### HIV GENES

As mentioned, a retrovirus minimally contains only the *gag* (group *a*ntigen gene), *pol* (polymerase), and *env* (envelope) genes. The viral messenger RNA is translated starting with *gag* (fig. 3). There is a translation termination signal at the end of the *gag* gene that is occasionally read through, resulting in a *gag-pol* protein. The *env* gene is translated only after the viral RNA is spliced to remove the *gag-pol* region. The protein products of all three genes are further modified by cleavage and other changes (phosphorylation and glycosylation), resulting in core virion proteins from *gag,* reverse transcriptase, protease, and integrase from *pol,* and envelope glycoproteins from *env.*

The HIV retrovirus is especially complicated. Not only does it have the *gag, pol,* and *env* genes, but it also has six other genes (fig. 4), including two main regulatory genes, *tat* and *rev.* One, *tat* (for *t*rans-*a*ctivating *t*ranscription factor), has a protein product that binds at a sequence

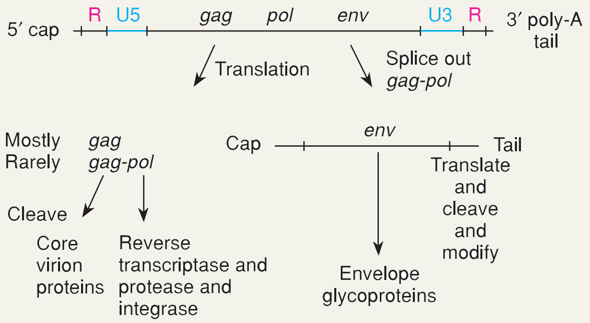

5′ cap
R
U5
gag
pol
env
U3
R
3′ poly-A tail
Translation
Splice out gag-pol
Mostly
Rarely
gag
gag-pol
Cleave
Core virion proteins
Reverse transcriptase and protease and integrase
Cap
env
Tail
Translate and cleave and modify
Envelope glycoproteins

**Figure 3** Expression of a retroviral mRNA. Translation begins with the *gag* gene and occasionally, due to read-through, proceeds through the *gag-pol* genes. The results are core virion proteins and the enzymes reverse transcriptase, protease, and integrase. Splicing must take place before *env* can be translated. Cleavage of the primary transcript and some modification produces envelope glycoproteins.

*continued*

## BOX 16.3 CONTINUED

in the long terminal repeat named TAR, for *trans*-*a*ctivating *r*esponse element. Tat enhances the processivity of transcription of the proviral DNA and also recruits chromatin-remodeling proteins to the promoter. The product of the other regulatory gene, *rev* (*r*egulation of *e*xpression of *v*irion proteins), binds at a region in the *env* gene called RRE (for *rev* *re*sponse *e*lement) and enhances the transport of viral messenger RNAs into the cytoplasm. Together, *tat* and *rev* are responsible for the major expression of viral structural genes (*gag, pol,* and *env*).

The four remaining genes—*vif, vpr, nef,* and *vpu*—are called the accessory genes because it first seemed that their action was not necessary for viral functioning. We now know that each gene produces a protein that has a role in viral replication and infectivity. The Vpr protein (*v*iral *p*rotein *R*) is involved in transporting the viral RNA to the nucleus. Vpr can also induce cell cycle arrest at G2, which may have a role in protecting infected cells from cytotoxic T-cell activities. Vpu (*v*iral *p*rotein *U*) degrades CD4; this action frees viral surface protein precursors from the endoplasmic reticulum. In addition, degradation of CD4 helps prevent superinfection of cells, keeping them alive longer. The main function of Vif (*v*iral *i*nfectivity *f*actor) is to stabilize the virion. Nef (*ne*gative *f*actor) was originally thought to be a negative regulator of viral activity, hence its name. However, it is now known that Nef can reduce production of cellular CD4 protein and enhance infection by viruses free in the blood.

### TESTING AND TREATMENT

AIDS testing is done by various techniques, such as western blots, looking for antibodies to the AIDS proteins, usually gp120, gp41, and reverse transcriptase. Initially, dideoxy nucleotides, such as the drug 3′-azido-2′, 3′-dideoxythymidine (AZT, fig. 5) and dideoxyinosine were used to treat AIDS. AZT is a thymidine analogue without a 3′-OH group, meaning that it causes chain termination during DNA replication. It seems that during the reverse transcription process, reverse transcriptase preferentially chooses AZT over normal thymidine-containing nucleotides, whereas mammalian DNA polymerases prefer the opposite. Thus, AZT preferentially prevents the reverse transcription of the HIV RNA, keeping it at levels that are not toxic to the cell.

Dideoxyinosine has the same effect and has also been licensed as an AIDS treatment. Unfortunately, the AIDS virus mutates at a high rate, rendering these single-substance treatments ultimately ineffective. In 1996, treatment success improved remarkably when new therapies involving combinations of drugs, including protease inhibitors, were developed. (Dr. David Ho of the Aaron Diamond AIDS Research Center in New York City was named *Time* Magazine's Man of the Year for his role in this therapy.) Thus, at the moment, optimism is rising that AIDS may be controllable and eventually curable.

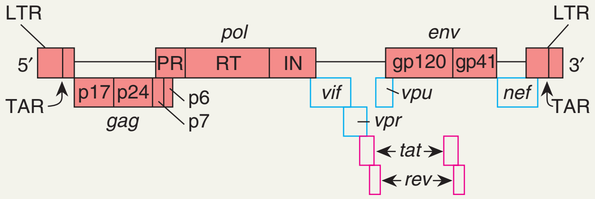


**Figure 4** The genome of HIV-1. Boxes represent different genes. The *gag* gene is responsible for four proteins, p17 (matrix), p24 (capsid), p7 (nucleocapsid), and p6. The *pol* gene is responsible for protease (PR), reverse transcriptase (RT), and integrase (IN). The *env* gene is responsible for two envelope proteins, gp120 and gp41. Intervening DNA separates the *tat* and *rev* genes into two parts each. TAR is in the long terminal repeat (LTR), and RRE is in *env.* (From R. H. Miller and N. Sarvar, "HIV Accessory Proteins as Therapeutic Targets," *Nature Medicine* 3:389–91, 1987. Copyright © 1987 Nature Publishing Group.)

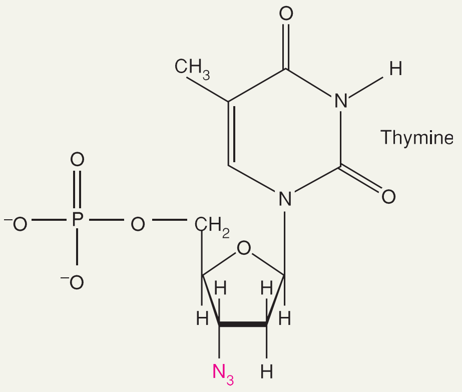


**Figure 5** AZT; it differs from deoxythymidine monophosphate at the 3′ position of the sugar.

## SUMMARY

**STUDY OBJECTIVE 1:** To examine the control of transcription in eukaryotes 465–469

In eukaryotes, specific transcription factors that can gain access to promoters generally control transcription. In addition, nucleosome structure can obstruct the RNA polymerase holoenzyme's access to the promoter. Signal transduction pathways provide environmental cues that lead to remodeled nucleosomes and the appearance of specific transcription factors at gene promoters that are scheduled to begin transcription. Methylation is involved in the control of gene expression in some higher eukaryotes; the methylation level is high in nontranscribed genes in the cells of these organisms. Transposons can also control gene expression. Mutable loci in corn and mating type in yeast are both determined by transposition. Z DNA may also play a role in the eukaryotic control of transcription.

**STUDY OBJECTIVE 2:** To analyze the genetic control of development in eukaryotes 469–484

The ultimate goal of the developmental geneticist is to understand the role of genes in controlling development, the orderly sequence of changes that give rise to a complex organism. Development does not have to proceed by permanently changing the chromosomes; cloning has shown that differentiated nuclei can be totipotent. The *Drosophila* embryo begins development with morphogens, diffusable messenger RNAs and proteins secreted by maternal-effect genes. These genes provide anterior, posterior, dorsoventral, and terminal patterns of transcription for zygotic segmentation genes; these segmentation genes fall into the gap, pair-rule, and segment-polarity classes. They eventually determine differential gene expression in neighboring cells. Finally, homeotic genes determine the fates of entire regions of the body. Flowering in angiosperms also involves repeated units (whorls) and homeobox genes.

**STUDY OBJECTIVE 3:** To study the mechanisms causing cancer 484–492

Cancer is a generic term for genetic diseases in which cells proliferate inappropriately. Mutations in oncogenes (cancer-causing genes) or tumor-suppressor genes, such as *p53,* can lead to cancer. Mutation can take place by base pair changes, chromosomal rearrangements, and amplification. Viruses can also bring oncogenes into cells, causing the activation of the oncogenes. For the full development of cancer, several genes usually must mutate.

**STUDY OBJECTIVE 4:** To study the genetic mechanisms that generate antibody diversity 492–504

Immunoglobulins (antibodies) have tremendous diversity; about $10^9$ different antibodies can be generated by the human genome. This number comes about by V(D)J joining between several genes among hundreds, as well as by junctional diversity caused by the location of the crossover points during site-specific recombination, template-free addition of codons (N segments), and somatic hypermutation. Similar diversity exists in T-cell receptors that recognize the major histocompatibility complex (MHC) proteins. These proteins present foreign polypeptides for destruction by the immune system.

## SOLVED PROBLEMS

**PROBLEM 1:** Relate the homeo box, homeo domain, and master-switch concepts.

***Answer:*** A master-switch gene is a gene in a eukaryote that controls many other genes. In a prokaryote, this control is achieved with operon organization; that is, many genes controlling the same function are transcribed as a unit. Thus, a gene that represses transcription of an operon represses all of the genes in that operon. A master-switch gene is viewed in a similar manner, given that polygenic transcripts are very rare in eukaryotes. A master-switch gene would translate to a specific transcription factor, a protein that might control transcription of many genes (a synexpression group). For this to happen, the master-switch gene would need to interact with DNA. Thus, the finding of a homeo box that transcribes a homeo domain in genes that control large phenotypic changes is consistent with this view. The homeo domain is the part of the transcription factor that binds to DNA.

**PROBLEM 2:** What stages in the formation of an immunoglobulin molecule generate diversity?

***Answer:*** Variability is generated through four general processes: choice of which subunit genes to combine, choice of how to combine these subunit genes, *de novo* generation of diversity at junctions, and unusually high mutation rates. Thus, in our description of the formation of a κ chain, diversity is added by (1) the choice of which variable and joining genes to combine; (2) recombinational variability at the point of recombination; (3) the creation of N segments at the junctions; and (4) somatic hypermutation.

**PROBLEM 3:** How can you reconcile the viral and mutational natures of cancer?

***Answer:*** The two theories are reconciled because both define cancer as a disease caused by the inappropriate actions of genes. In the mutational view, inappropriate activity is generated by a gene mutation. In the viral view, a gene brought into the cell by a virus generates the inappropriate activity.

## EXERCISES AND PROBLEMS*

### CONTROL OF TRANSCRIPTION IN EUKARYOTES

1. Diagram the sequence on the yeast third chromosome as the mating type changes from **a** to α and back again.
2. What are the differences between a general transcription factor and a specific transcription factor?
3. Tissue culture cells are exposed for five minutes to radioactive dUTP in the presence or absence of 5-azacytidine. Radioactivity in RNA is determined to be 1,500 counts per minute without azacytidine and 27,300 in the presence of azacytidine. Propose an explanation to account for these results.
4. A retrovirus, lacking a cellular oncogene, is shown to be integrated 3 kilobases from a proto-oncogene. When the RNA for this oncogene is quantified, infected cells are found to have ten times more oncogene-specific messenger RNA than uninfected cells. How can you account for this increase in RNA synthesis?

### PATTERNS IN DEVELOPMENT

5. What is *genomic equivalence,* and why is explaining it a central problem in developmental genetics?
6. What is the relationship between parasegments and segments in the developing *Drosophila* embryo?
7. What are the three classes of segmentation genes in *Drosophila* embryos? What are the effects of mutations of genes in each class?
8. How does the *hunchback* gene function in *Drosophila* development?
9. What are the differences between a syncitial and a cellular blastoderm in a *Drosophila* embryo?
10. What is meant by the statement that homeotic genes have been conserved evolutionarily?
11. What are the four regions of the body plan of the developing *Drosophila* embryo laid out by maternal-effect genes? What are the four major maternal-effect genes?
12. What is a morphogen? How does the Bicoid protein of *Drosophila* function as one?
13. What is the helix-turn-helix motif of DNA binding? What other motifs are known for DNA-binding proteins?
14. If drugs that inhibit transcription are injected into fertilized eggs, early cell division and protein synthesis still occur. Why?
15. Why do you suppose so much early research on developmental genetics was done with amphibians?

### CANCER

16. What gross chromosomal abnormalities are associated with cancers?
17. From the pedigree of figure 16.26, what modes of inheritance would be consistent with each type of cancer, assuming that a single gene controlled each?
18. What chromosomal abnormality is associated with retinoblastoma? with Wilm's tumor?
19. What is the proposed mechanism of action of the retinoblastoma gene? What evidence supports this mechanism? Why is it called an anti-oncogene?
20. Retinoblastoma has been called a recessive oncogene. Explain.
21. What are the general forms of animal viruses? What types of genetic material do they have?
22. What is the minimal genetic complement of a retrovirus? What does each of the genes code for?

---

*Answers to selected exercises and problems are on page A-19.

**23.** What translation mechanisms exist for the expression of the genes of a retrovirus?

**24.** Assume that a particular oncogene produces a growth factor.
   **a.** How could a retrovirus affect the oncogene so that the cell becomes cancerous?
   **b.** How could you test your hypothesis?

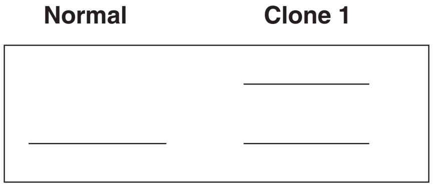
**25.** What are the differences among v-*src*, c-*src*, and proto-*src* genes?

**26.** How can the proto-*src* gene be activated?

**27.** What is the evidence that the c-*src* gene came before the v-*src* gene?

**28.** How does translocation activate the c-*myc* gene in Burkitt's lymphoma?

**29.** A cDNA probe for a proto-oncogene is constructed. Cellular DNA from normal cells and a clone of cells infected with a retrovirus that lacks the oncogene (clone 1) is digested with a particular restriction enzyme. The DNA is separated in a gel and hybridized with the radioactive probe. The results appear in the following figure.



Interpret these results by describing where the retrovirus has inserted.

### IMMUNOGENETICS

**30.** What is the general mechanism that allows an antibody to "recognize" an antigen?

**31.** What components go into making an Ig light chain? a heavy chain?

**32.** How many different antibodies does a B lymphocyte produce? How many can it potentially produce before it differentiates?

**33.** What are the nucleotide recognition signals in V-J joining?

**34.** What are B and T lymphocytes? What roles do they play in the immune response?

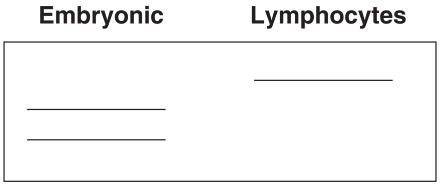
**35.** What is a T-cell receptor?

**36.** What is the major histocompatibility complex?

**37.** A disorder of the immune system is characterized by a complete lack of antibody production. Provide two possible molecular defects that would result in such a condition.

**38.** Many alleles for the genes for the constant region of antibodies have been found. Suppose that two such alleles for the $\lambda$ light chain are called $c_1$ and $c_2$. In a heterozygote, $c_1c_2$, some cells are found to make only $c_1$ and others only $c_2$. Propose an explanation.

**39.** Complementary DNA is made from messenger RNA for the light chain of an antibody molecule. DNA from embryonic cells and from mature B lymphocytes is isolated and digested with a restriction enzyme, and the fragments are separated in a gel. Radioactive cDNA is used to probe this gel, and the results appear in the figure that follows. Provide an explanation for these results.



## CRITICAL THINKING QUESTIONS

**1.** The *E1B* gene of adenovirus produces a protein that binds with p53, allowing the virus to multiply in the cell. Given that more than 50% of cancer cells lack p53 activity, how might you engineer the adenovirus to attack only cells without p53 activity? That is, can you engineer adenovirus to attack a large proportion of cancerous cells?

**2.** Given what you know about flower development in plants, what might be the simplest mechanism plants could use to produce male-only or female-only flowers?

*Suggested Readings for chapter 16 are on page B-15.*

# 17

# NON-MENDELIAN INHERITANCE

## STUDY OBJECTIVES

1. To analyze the inheritance patterns of maternal effects 509
2. To analyze the patterns of cytoplasmic inheritance 511
3. To analyze the patterns of imprinting 524

## STUDY OUTLINE

**Determining Non-Mendelian Inheritance** 509
**Maternal Effects** 509
  Snail Coiling 509
  Moth Pigmentation 510
**Cytoplasmic Inheritance** 511
  Mitochondria 511
  Chloroplasts 515
  Infective Particles 518
  Prokaryotic Plasmids 522
**Imprinting** 524
*Summary* 524
*Solved Problems* 525
*Exercises and Problems* 525
*Critical Thinking Questions* 527

Artificially colored scanning electron micrograph of a mitochondrion in the cytoplasm of an intestinal epithelial cell. (© Professors P. Motta & T. Naguro/ SPL/Photo Researchers, Inc.)

The phenotype can be controlled by chromosomal genes behaving according to Mendelian rules and the environment. In this chapter, we deal with another mode of inheritance, non-Mendelian inheritance (also called extrachromosomal, cytoplasmic, and nonchromosomal inheritance; maternal effects; or imprinting). **Maternal effects** are the influences of a mother's genotype on the phenotype of her offspring; examples include snail coiling and moth pigmentation (we started a discussion of maternal effects in chapter 16, when we looked at development in *Drosophila*). **Cytoplasmic inheritance** is controlled by nonnuclear genomes found in chloroplasts, mitochondria, infective agents, and plasmids. And *imprinting* is a process in which gene expression depends on the parent from which the gene came. None of these modes of inheritance follow the usual Mendelian rules and ratios.

Maternal effects result from the asymmetric contribution of the female parent to the development of zygotes. Although both male and female parents contribute equally to the zygote in terms of chromosomal genes (with the exception of sex chromosomes), the sperm rarely contributes anything to development other than chromosomes. The female parent usually contributes the zygote's initial cytoplasm and organelles. Zygotic development, therefore, usually begins within a maternal milieu, so that the maternal cytoplasm directly affects zygotic development (see chapter 16).

Cytoplasmic inheritance refers to the inheritance pattern of organelles and parasitic or symbiotic particles that have their own genetic material. Chloroplasts, mitochondria, bacteria, viruses, and, of course, plasmids all have their own genetic material. These genomes are open to mutation. As we shall see, their inheritance pattern does not follow Mendel's rules for chromosomal genes.

Imprinting occurs in more than twenty genes and is responsible for several human diseases.

## DETERMINING NON-MENDELIAN INHERITANCE

How does one determine that a trait is inherited? The question does not have as obvious an answer as we might expect. Environmentally induced traits can mimic inherited phenotypes, as with the phenocopies we discussed in chapter 5. For example, the inheritance of vitamin D-resistant rickets is mimicked by lack of vitamin D in the diet. It is possible to determine that the rickets is not inherited by simply administering adequate quantities of vitamin D. Inherited rickets does not respond to vitamin D until about 150 times the normally adequate amount is administered.

Some environmentally induced traits persist for several generations. For example, a particular *Drosophila* strain that normally grows at 21° C was exposed to 36° C for twenty-two hours. Dwarf progeny were produced. When they were mated among themselves, fewer and fewer dwarfs appeared in each generation, but smaller-than-normal flies were produced as late as the fifth generation. The appearance of an environmentally induced trait that persists for several generations has been termed **dauermodification.**

Extrachromosomal inheritance is usually identified by the odd results of reciprocal crosses. If the progeny of reciprocal crosses are not followed for several generations, the results can be misleading when extrachromosomal inheritance is involved. Where feasible, nuclear transplantation has proved useful in identifying extrachromosomal inheritance. In this technique, the nucleus of a cell, such as an amoeba or frog egg, is removed by microsurgery or destroyed by radiation, and another nucleus substituted. Thus, not only can a nucleus be isolated from its cytoplasm, but various nuclei can be implanted in the same cytoplasm.

A similar experiment, called a *heterokaryon test,* can be done with various fungi such as *Neurospora* and *Aspergillus:* Mycelia can fuse and form a heterokaryon, a cell containing nuclei from different strains. Thus, nuclei of both strains exist in the mixed cytoplasm. Subsequently, spores (conidia) that have one or the other nucleus in the mixed cytoplasm can be isolated. The phenotype of the colonies produced from these isolated conidia show whether the trait under observation is controlled by the nucleus or the cytoplasm.

Chromosomal genes in a particular cytoplasm can also be isolated by repeated backcrossing of offspring with the male-parent type. In each cross, the content of the female chromosomal genes is halved, but, presumably, the cytoplasm remains similar to the female line. Thus, after several generations, male genes can be isolated in female cytoplasm. The phenotypic results of the final cross will indicate whether inheritance was chromosomal or extrachromosomal.

## MATERNAL EFFECTS

### *Snail Coiling*

Snails are coiled either to the right (dextrally) or to the left (sinistrally) as determined by holding the snail with the apex up and looking at the opening. The snail is dextrally coiled if the opening comes from the right-hand side and sinistrally coiled if it comes from the left-hand side (fig. 17.1). The inheritance pattern of the coiling is at first perplexing.

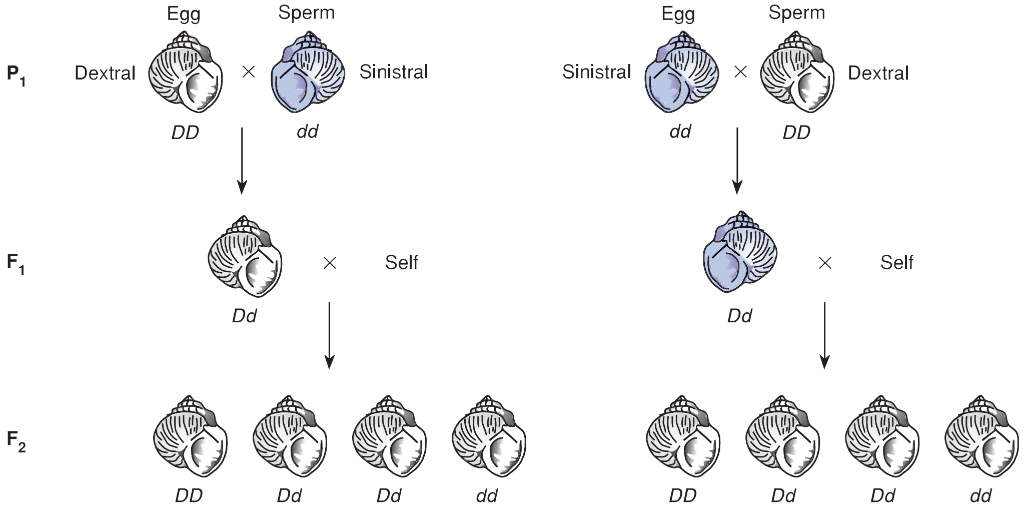


**Figure 17.1** Inheritance of coiling in the pond snail, *Limnaea peregra.* Reciprocal crosses (*D,* dominant dextral; *d,* recessive sinistral coiling) are shown (*DD* mated with *dd* in each case). The $F_1$ individuals in both crosses have the *Dd* genotype but reflect the mother's genotype in respect to coiling; *DD* mothers produce dextrally coiled offspring, whereas *dd* mothers produce sinistrally coiled offspring. The $F_2$ individuals in both cases are identical because the genotypes of the $F_1$ mothers are identical (*Dd*). The coiling of a snail's shell is determined by its mother's genotype, not phenotype.

In the left half of figure 17.1, a dextral snail provides the eggs, and a sinistral snail provides the sperm. The offspring are all dextral; presumably, therefore, dextral coiling is dominant. When the $F_1$ are self-fertilized (snails are hermaphroditic), all the offspring are dextrally coiled. The result is unexpected. Nevertheless, when the $F_2$ are self-fertilized, one-fourth produce only sinistral offspring, and three-fourths produce only dextral offspring. If self-fertilization is continued through ensuing generations, this 3:1 phenotypic ratio will be revealed as a Mendelian 1:2:1 genotypic ratio, thereby reaffirming the notion of a single locus with two alleles, and dextral dominant. However, something interfered with the expected phenotypic pattern.

When the reciprocal cross is made (fig. 17.1, *right*), the $F_1$ have the same genotype as just described but are coiled sinistrally, as is the female parent. From here on, the results are exactly the same for both crosses. In both cases, the $F_1$ are phenotypically similar to the female parent even though the offspring in both crosses have the same genotype (*Dd*). The explanation is that the genotype of the maternal parent determines the phenotype of the offspring, with dextral dominant. Thus, the *DD* mother in figure 17.1 produces $F_1$ progeny that are dextral with a *Dd* genotype, and the *dd* mother produces progeny that are sinistral with the same *Dd* genotype. Why does this pattern occur?

A process of **spiral cleavage** takes place in the zygotes of mollusks and some other invertebrates. The spindle at mitosis is tipped in relation to the axis of the egg. If the spindle is tipped one way, a snail will be coiled sinistrally; if it is tipped the other way, the snail will be coiled dextrally. The direction of tipping is determined by the maternal cytoplasm, which is under the control of the maternal genotype. Obviously, maternal control affects only one generation—in each generation, the coiling is dependent on the maternal genotype.

## Moth Pigmentation

There are other examples of maternal effects in which the cytoplasm of the mother, under the control of chromosomal genes, controls the phenotype of her offspring. In the flour moth, *Ephestia kühniella,* kynurenin, which is a precursor for pigment, accumulates in the eggs. The recessive allele, *a,* when homozygous, results in a lack of kynurenin. Reciprocal crosses give different results for larvae and adults. When a nonpigmented female is crossed with a pigmented male, the results are strictly Mendelian; but when the mother is pigmented ($a^+a$), all the larvae are pigmented regardless of genotype (fig. 17.2). The initial larval pigmentation comes from residual kynurenin in the eggs, which is then diluted out so that an adult's pigmentation conforms to its own genotype.

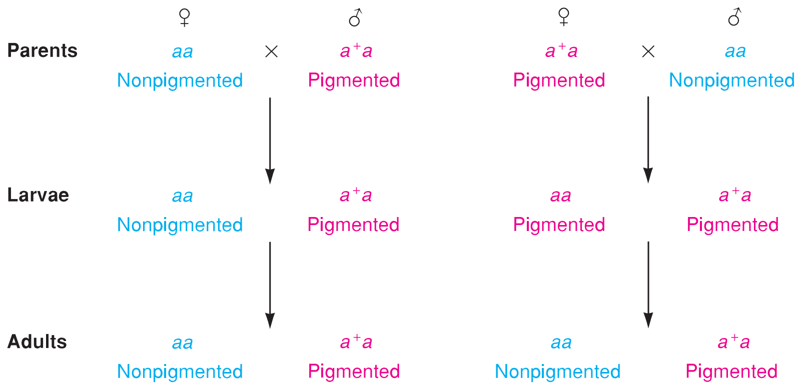


**Figure 17.2** Inheritance pattern of larval and adult pigmentation in the flour moth, *Ephestia kühniella.* A single locus controls the presence ($a^+$) or absence ($a$) of kynurenin. In the cross on the *left,* the mother is $aa$ (nonpigmented). Her $aa$ offspring, in both the larval and adult stages, are also nonpigmented. In the reciprocal cross *(right),* the mother has the $a^+a$ genotype and is pigmented. Her $aa$ offspring are nonpigmented as adults but are pigmented as larvae because of residual kynurenin from the egg, which eventually dilutes out.

## CYTOPLASMIC INHERITANCE

### *Mitochondria*

The **mitochondrion** is an organelle in eukaryotic cells in which the electron transport chain takes place. The actual number of mitochondria per cell can be determined by serial sectioning of whole cells and examination under the electron microscope. This is a tedious and difficult procedure. Estimates range between ten and ten thousand per cell, depending on the organism and cell type. As far as we are concerned, the most interesting aspect of the mitochondrion is that it has its own DNA. In most animal cells, the mitochondrial DNA (mtDNA) is a circle of about sixteen thousand base pairs (fig. 17.3). However, some organisms (yeast, higher plants) have mitochondrial DNAs five to twenty-five or more times larger than in animals. And some organisms have linear mitochondrial chromosomes.

Two general patterns are found in mitochondrial inheritance in animals. First, the mitochondria are generally inherited in a maternal fashion; that is, the male gamete usually does not contribute mitochondria to the zygote. However, a small amount of "leakiness" occurs in this process. For example, it has recently been shown that about one mitochondrion per thousand is of paternal origin in mice. In some species, such as mussels, it appears that mitochondrial inheritance is biparental. That is, the population of mitochondria in an offspring derives almost equally from the male and female parent. In some gymnosperm plants, such as coastal redwoods, mitochondria are inherited paternally—only paternal mitochondria are passed into the zygote. However, these are all exceptions to the general rule of maternal inheritance of mitochondria.

The second general pattern of mitochondrial inheritance is **homoplasmy,** the existence of a uniform population of mitochondria within an organism. In general, all the mitochondria within an individual are genetically identical. Certainly, biparental inheritance and leakiness of paternal mitochondria violate that principle, resulting in **heteroplasmy,** a heterogeneity of mitochondria within a cell or organism.

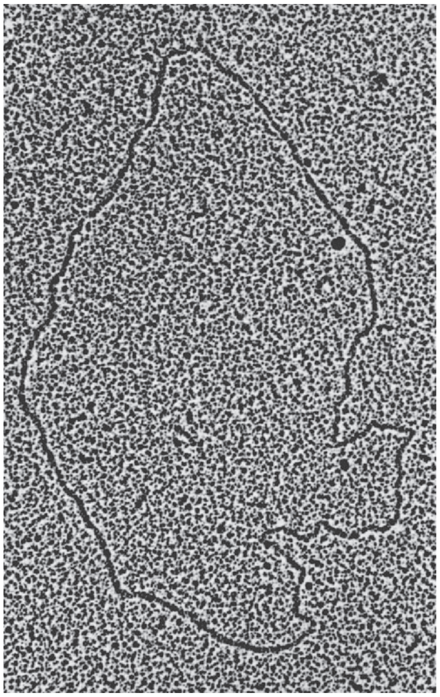


**Figure 17.3** Electron micrograph of the circular DNA from within a mouse cell mitochondrion. Magnification 48,000×. (M. M. K. Nass, "The circularity of mitochondrial DNA," *Proceedings of the National Academy of Sciences, USA,* 56 (1966):1215–22. Reproduced by permission of the author.)

### *Mitochondrial Genomes*

Numerous mitochondrial DNAs have been sequenced, including the human mitochondrial DNA, which is 16,569 base pairs long. It is a model of economy, with very few noncoding regions and no introns (fig. 17.4). Each strand of the duplex is transcribed into a single RNA product that is then cut into smaller pieces, primarily by freeing the twenty-two transfer RNAs interspersed throughout the genome. Also formed are a 16S and a 12S ribosomal RNA. Although proteins and small molecules such as ATP and tRNAs can move in and out of the mitochondrion, large RNAs cannot. Thus, the mitochondrion must be relatively self-sufficient in terms of the RNAs needed for protein synthesis. We previously discussed mitochondrial protein synthesis when we looked at unique attributes of the mitochondrial genetic code in chapter 11.

Oxidative phosphorylation, the process that occurs within the mitochondrion, requires at least sixty-nine polypeptides. The human mitochondrion has the genes for thirteen of these: cytochrome b, two subunits of ATPase, three subunits of cytochrome-c oxidase, and seven subunits of NADH dehydrogenase. The remaining polypeptides needed for oxidative phosphorylation are transported into the mitochondrion; they are synthesized in the cytoplasm under the control of nuclear genes. Proteins targeted for entry into the mitochondrion have special signal sequences (see chapter 11).

The signal sequences range up to eighty-five amino acids long. Signal sequences examined so far do not have consensus amino acids but do have certain attributes (fig. 17.5), including a somewhat regular alternation of basic (positively charged) and hydrophobic (negatively charged) residues. In addition, they form α helices with opposite hydrophobic and hydrophilic faces that must somehow be important in the protein's ability to enter the mitochondrion. When a signal sequence (such as that in fig. 17.5) is attached to nonmitochondrial proteins by DNA manipulations, those proteins are transported into the mitochondrion.

The mitochondrial ribosomal RNA is more similar to prokaryotic ribosomal RNA than to eukaryotic ribosomal RNA. The mitochondrial ribosome, although constructed of imported cellular proteins, is sensitive to prokaryotic antibiotics; for example, streptomycin and chloramphenicol inhibit their function. This affinity (close resemblance) between mitochondria and prokaryotes is strong support for the symbiotic origin of mitochondria. That is, we now accept the model advocated by L. Margulis that organelles such as mitochondria and chloroplasts were originally free-living bacteria and cyanobacteria, respectively. These prokaryotes invaded or were eaten by early cells and, over evolutionary time, became the organelles we see today. Since they arose as prokaryotes, these organelles retain certain evolutionary similarities to other prokaryotes.

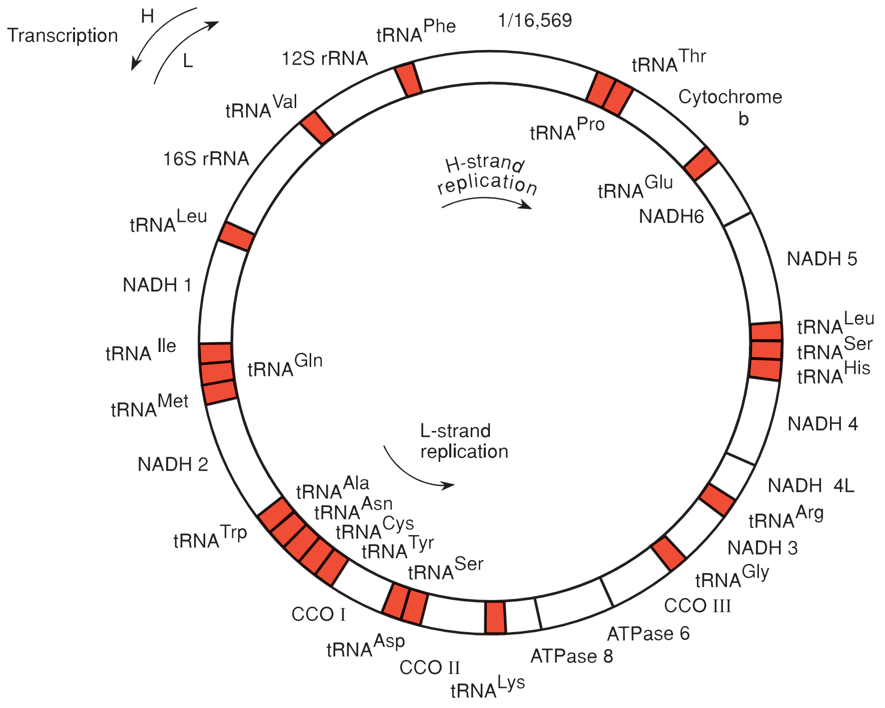


**Figure 17.4** Gene map of the human mitochondrial chromosome. All but nine loci are on the heavy (H) strand. The light-strand (L) loci are labeled inside the circle; the H-strand loci are labeled on the outside. Also shown are the origins of H- and L-strand replication and the directions of transcription. The twenty-two tRNA genes are colored *red.* NADH refers to NADH dehydrogenase (subunits 1–4, 4L, 5, and 6); CCO refers to cytochrome-c oxidase (subunits I–III). (Source: Data from V. McKusick, *Mendelian Inheritance in Man,* 7th edition, 1986.)

Lynn Margulis (1938– ).
(Courtesy of Lynn Margulis, Boston University Photo Services.)

Among the mitochondrial DNAs that have been sequenced from different organisms, we see great variation in content and organization. Yeast mitochondrial DNA, for example, is not as economical as human mitochondrial DNA. Yeast mitochondrial DNA, about five times larger than human mitochondrial DNA, has noncoding regions as well as introns. Because mitochondria are similar in structure and biochemistry to prokaryotic cells, given the general lack of introns in prokaryotic genes, it was surprising to find introns in yeast mitochondrial DNA. These genes most probably arose later as nuclear genes that were then "captured" by the mitochondria, possibly by recombination with nuclear DNA.

Of the many mitochondria sequenced to date (about 175 at the beginning of 2001), the sizes range from less than 6 to more than 200 kilobases and from 3 to 97 genes. With this wide range of genes present, the only generality we can make about mitochondrial DNA is that the large and small segments of the mitochondrial ribosomal RNA, as well as most of the mitochondria's transfer RNAs, are usually coded by the mitochondria's own genome, as are several proteins in respiratory complexes III and IV (cytochrome c oxidase and cytochrome c oxidoreductase). Once the interaction within the mitochondrial-nuclear genetic system is clearly understood, we might expect to see several different inheritance patterns—following either cytoplasmic or nuclear lines—for the genetic defects that lead to interruption of cellular respiration. Among the best-studied phenotypes with such inheritance patterns are the *petite* mutations of yeast.

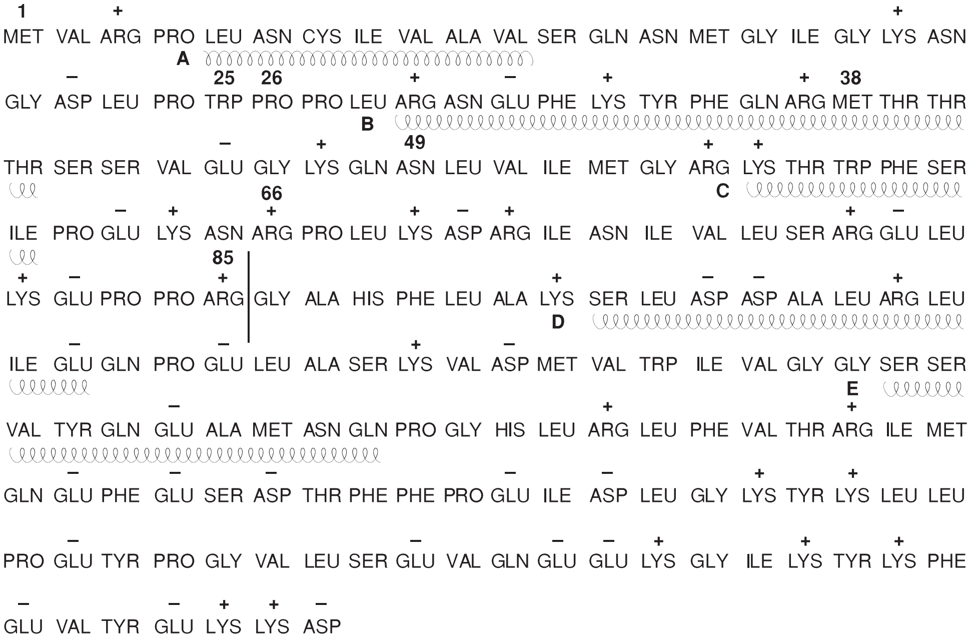

1
MET VAL ARG PRO LEU ASN CYS ILE VAL ALA VAL SER GLN ASN MET GLY ILE GLY LYS ASN
A
25 26 38
GLY ASP LEU PRO TRP PRO PRO LEU ARG ASN GLU PHE LYS TYR PHE GLN ARG MET THR THR
B
49
THR SER SER VAL GLU GLY LYS GLN ASN LEU VAL ILE MET GLY ARG LYS THR TRP PHE SER
C
66
ILE PRO GLU LYS ASN ARG PRO LEU LYS ASP ARG ILE ASN ILE VAL LEU SER ARG GLU LEU
85
LYS GLU PRO PRO ARG GLY ALA HIS PHE LEU ALA LYS SER LEU ASP ASP ALA LEU ARG LEU
D
ILE GLU GLN PRO GLU LEU ALA SER LYS VAL ASP MET VAL TRP ILE VAL GLY GLY SER SER
E
VAL TYR GLN GLU ALA MET ASN GLN PRO GLY HIS LEU ARG LEU PHE VAL THR ARG ILE MET
GLN GLU PHE GLU SER ASP THR PHE PHE PRO GLU ILE ASP LEU GLY LYS TYR LYS LEU LEU
PRO GLU TYR PRO GLY VAL LEU SER GLU VAL GLN GLU GLU LYS GLY ILE LYS TYR LYS PHE
GLU VAL TYR GLU LYS LYS ASP
+
−

**Figure 17.5** The amino acid sequence of mouse dihydrofolate reductase. Numbers refer to sequential amino acids. The first eighty-five amino acids serve as the signal sequence for transport into mitochondria. Five α-helical regions exist in the protein (A–E). Positively and negatively charged amino acids are marked with (+) and (−) signs. (Reprinted by permission from E. C. Hurt and G. Schatz, "A cytosolic protein contains a cryptic mitochondrial targeting signal," *Nature,* Volume 325, p. 499, 1987. Copyright © 1987 Macmillan Magazines, Ltd.)

### *Petites*

Under aerobic conditions, yeast grows with a distinctive colony morphology. Under anaerobic conditions, the colonies are smaller, and the structures of the mitochondria are reduced. Occasionally, when growing aerobically, small, anaerobiclike colonies appear; but in these colonies, the mitochondria appear perfectly normal. These colonies are caused by **petite mutations.** When petites are crossed with the wild-type, three modes of inheritance emerge (fig. 17.6). The *segregational petite,* caused by mutation of a chromosomal gene, exhibits Mendelian inheritance. The *neutral petite* is lost immediately upon crossing to the wild-type. The *suppressive petite* shows variability in expression from one strain to the next but is able to convert the wild-type mitochondria to the petite form. All petites represent failures of mitochondrial function, whether the function is controlled by the mitochondria themselves or by the cell's nucleus; they usually lack one or another cytochrome.

Although the mechanisms that produce neutral and suppressive petites are not known with certainty, their DNA has supplied some interesting information. In some petites, no change in the buoyant density of the DNA is found. (*Buoyant density,* a term that describes the position at which the DNA equilibrates during density-gradient centrifugation, is a measure of the composition of the molecule; see chapter 15.) In other petites, changes in buoyant density range from very small to the complete absence of DNA.

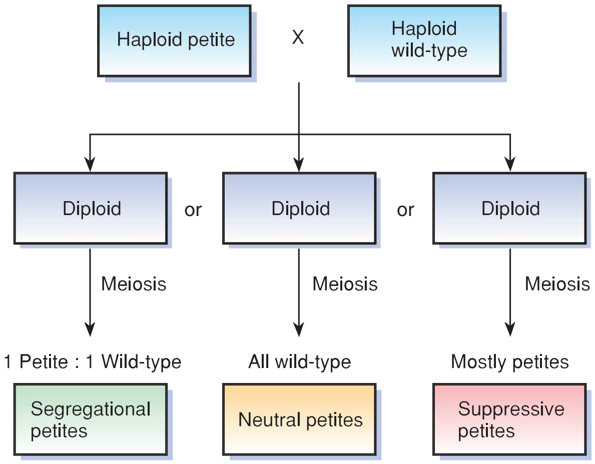


**Figure 17.6** Petite yeasts categorized on the basis of segregation patterns. Three types of petites are recognized (segregational, neutral, and suppressive), depending on the meiotic segregation pattern of petite × wild-type diploids. Segregational petite heterozygotes segregate a 1:1 ratio of spores; neutral petites are lost when heterozygous; and suppressive petites act in a dominant fashion under the same circumstances.

Petites, therefore, can be the result of an approximation to a point mutation (with no measurable change in the buoyant density of the DNA), marked changes in the DNA, or the total absence of DNA. In most petites, protein synthesis within the mitochondrion is lacking. Any and all of these changes produce the petite (anaerobic-like) phenotype.

Neutral petites seem to have mitochondria that entirely lack DNA. When neutral petites are crossed with the wild-type to form diploid cells, the normal mitochondria dominate. During meiosis, virtually every spore receives large numbers of normal mitochondria; the progeny are, therefore, all normal.

Suppressive petites could exert their influence over normal mitochondria in one of two ways. The suppressive mitochondria might simply out-compete the normal mitochondria and take over; they might simply reproduce faster within a cell. Alternatively, crossing over between the DNA of the suppressive petite and the wild-type might affect the normal DNA if the suppressive petite's DNA were severely damaged. Presumably, recombination in mitochondrial DNAs occurs when two or more mitochondria fuse, bringing the two different sets of DNA in contact within the same organelle. Recombination would presumably take place by normal crossover mechanisms.

If large portions of the DNA from the suppressive mitochondria were missing or altered, recombination with the normal mitochondria's DNA might exchange some of this damaged DNA. Several experiments have crossed a suppressive petite and a wild-type, each with mitochondrial DNA of known buoyant density. The DNAs of the offspring colonies, which were petites, were of various buoyant densities. For example, when a normal strain with mitochondrial DNA with a buoyant density of 1.684 g/cm$^3$ was crossed with a suppressive petite with a buoyant density of 1.677 g/cm$^3$, the offspring colonies' mitochondrial DNA had buoyant densities of 1.671, 1.674, and 1.683 g/cm$^3$. Such information supports the notion that the suppressive character takes over a colony by way of recombination.

### *Human Mitochondrial Inheritance*

In human beings, certain diseases trace their dysfunction to mitochondrial pathologies. The first such disease, Luft disease, characterized by excessive sweating and general weakness, was reported in 1962. In 1988, Douglas Wallace and his colleagues showed that *Leber optic atrophy* is a cytoplasmically inherited disease. This disease causes blindness, with a median age of onset of twenty to twenty-four years. The onset age and phenotype are variable, depending on the degree of heteroplasmy in the individual. Apparently, defects in mitochondria are not tolerable in the optic nerve, which demands a great deal of energy. The disease also does some damage to the heart. Pedigrees showed that Leber optic atrophy is transmitted only maternally. Sequencing of mitochondrial DNAs in affected families pinned down the disease to a point mutation, a change in nucleotide 11,778, which is in the gene for NADH dehydrogenase subunit 4 (see fig. 17.4). A guanine is changed to an adenine at codon 340, which converts an arginine to a histidine. This is the first human disease traced to a specific mitochondrial DNA mutation. Since 1962, over one hundred diseases, including some of the general symptoms of aging and cancer, have been attributed to mitochondrial pathology.

### Antibiotic Influences

Since the machinery of mitochondrial protein synthesis is prokaryotic in nature, antibiotics such as chloramphenicol and erythromycin can inhibit it. These antibiotics elicit a petite-type growth response in yeast. Antibiotic-resistant strains can be obtained by growing yeast on the antibiotic; only resistant mutants will grow. The resistance appears to be inherited in the mitochondrial, not the cellular, DNA. A mitochondrial inheritance pattern results, with crosses between a resistant and a sensitive (wild-type) yeast, as shown in figure 17.7. The resulting diploid colonies segregate both resistant and sensitive cells. Although not expected on the basis of a chromosomal gene, the random sorting of mitochondria through cell division could result in a wild-type cell containing only sensitive mitochondria. Since some yeast have only one to ten mitochondria per cell, this random assortment of sensitive mitochondria can be expected to occur at a relatively high rate.

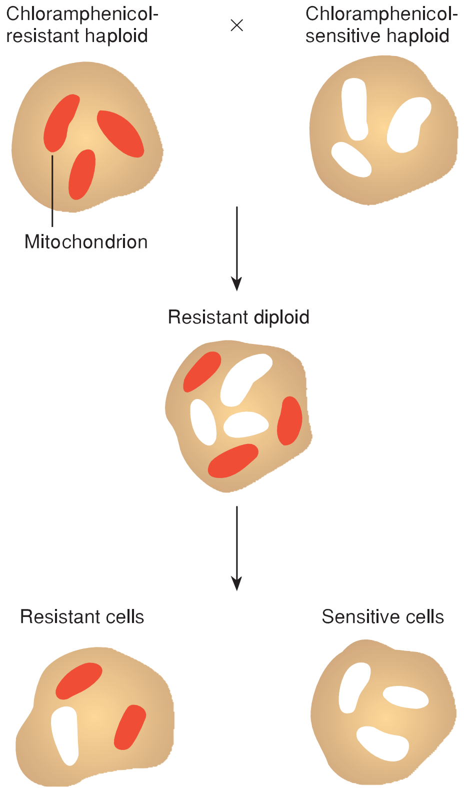

Chloramphenicol-resistant haploid
×
Chloramphenicol-sensitive haploid
Mitochondrion
Resistant diploid
Resistant cells
Sensitive cells

**Figure 17.7** Inheritance of antibiotic (chloramphenicol) resistance in yeast. Resistant and sensitive cells are produced by a diploid cell that resulted from a cross of resistant and sensitive haploids. The segregation is not in a simple Mendelian ratio, but depends on the random assortment of mitochondria. Sensitive cells have no resistant mitochondria. Resistant cells have resistant mitochondria.

## Chloroplasts

The **chloroplast** is the chlorophyll-containing organelle that carries out photosynthesis and starch-grain formation in plants (fig. 17.8). Chloroplasts are referred to as **plastids** before chlorophyll develops. However, when grown in the dark (and under some other circumstances), plastids do not develop into chloroplasts, but remain reduced in size and complexity. These undeveloped plastids, referred to as **proplastids,** are each about the size and shape of a mitochondrion.

Like mitochondria, chloroplasts contain DNA and ribosomes, both with prokaryotic affinities. The DNA of chloroplasts (cpDNA) is a circle that ranges in size from 85 kilobases (kb) in the green alga *Codium* to as large as 2,000 kilobases in the green alga *Acetabularia.* Thus, chloroplast DNA is minimally about five times the size of an animal mitochondrial DNA. The chloroplast DNA, like mitochondrial DNA, controls the production of transfer RNAs, ribosomal RNAs, and some of the proteins found within the organelle. From the more than nineteen chloroplast DNAs that have been sequenced, there seem to be about one hundred genes in the chloroplast genome. About thirty code for the subunits of the five photosynthetic protein complexes: photosystem I, photosystem II, ribulose bisphosphate carboxylase-oxygenase, cytochrome *b6-f* complex, and ATP synthase. About sixty genes code for the protein synthesis apparatus of the chloroplast. Scientists believe that the chloroplast evolved from symbiotic cyanobacteria (blue-green algae), which have many affinities with the chloroplast: The ribosomal RNA of cyanobacteria will hybridize with the DNA of chloroplasts.

The similarities between mitochondria and chloroplasts make it possible to predict the inheritance patterns of chloroplast mutations on the basis of existing

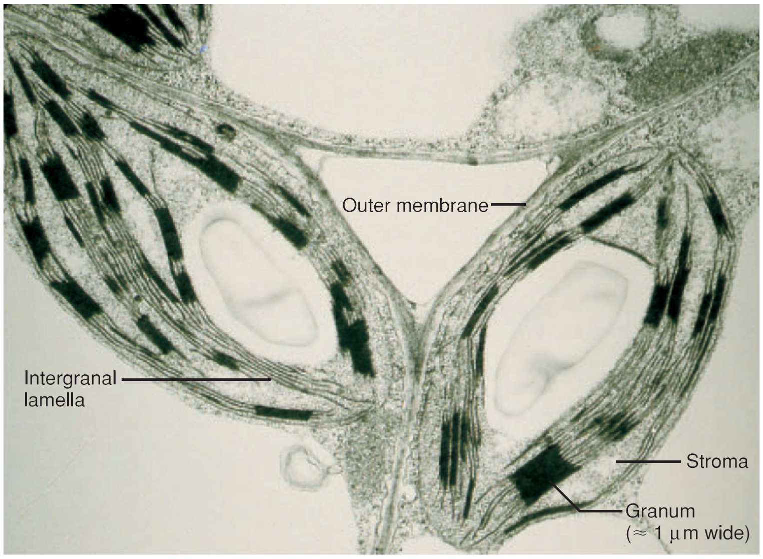


**Figure 17.8** Electron micrograph of lettuce chloroplasts. The chloroplast consists of an outer membrane, stacks of grana, lamellae, and stroma. Magnification 3,570×. (© Dr. J. Burgess/Science Photo Library/Photo Researchers, Inc.)

knowledge of mitochondrial genetics: We should find both chromosomal and plastid mutants of chloroplast functions. Simple segregation should occur in the chromosomal mutations, and cytoplasmic patterns of inheritance should occur with the chloroplast DNA mutations. Investigation of these inheritance patterns is complicated by the fact that plant cells have both mitochondria and chloroplasts. Since both have prokaryotic affinities, it is sometimes difficult to determine whether a genetic trait is due to a defect in the genetic system of the chloroplast or the mitochondrion. Like mitochondria, chloroplasts generally show homoplasmy and maternal inheritance, although, as in mitochondria, there are exceptions. For example, gymnosperms usually have paternal inheritance of chloroplasts.

Lesions in the photosystems of the chloroplast result in proplastid formation, with a loss of green color. When proplastid formation occurs in a particular tissue of a plant, variegation results. That is, there are both green and white parts, often as stripes. Some interesting genetic studies have focused on the inheritance of variegation, especially in the interaction of chloroplast and chromosomal genes.

## Zea mays

M. Rhoades worked on the variegation in corn (*Zea mays*) controlled by the *iojap* chromosomal locus, which, when homozygous, prevents proplastids from developing into chloroplasts and thus results in variegation. The *iojap*-affected plastids do not contain ribosomes or ribosomal RNA; they therefore lack protein synthesis.

Marcus M. Rhoades (1903–1991). (Courtesy of Indiana University Office of Communications and Marketing.)

The interaction of chromosomal and extrachromosomal inheritance is shown in the reciprocal crosses depicted in figure 17.9. One cross produces results exactly as would be predicted on the basis of simple Mendelian inheritance, with the homozygous recessive genotype (*ijij*) inducing variegation. When the reciprocal cross is carried out, blotch variegation is seen in both the $F_1$ and $F_2$ that carry the dominant *Ij* allele.

This inheritance pattern is caused by the fact that the pollen grain in corn does not carry any chloroplasts, whereas the ovule does. Thus, the first cross in figure 17.9 deals with the passage of normal chloroplasts only into the $F_2$ generation. In the $F_2$, the *ijij* genotype then induces variegation. The chloroplasts of the pollen parent are unimportant because they do not enter the $F_1$. In the reciprocal cross, however, because the stigma parent is variegated, the $F_1$ is heterozygous but carries proplastids from the ovule that remain proplastids even under the dominant normal (*Ij*) allele. Therefore, regions of colorless cells produce white spots (blotchy variegation). Once the *ij* allele induces chloroplasts to become pro-

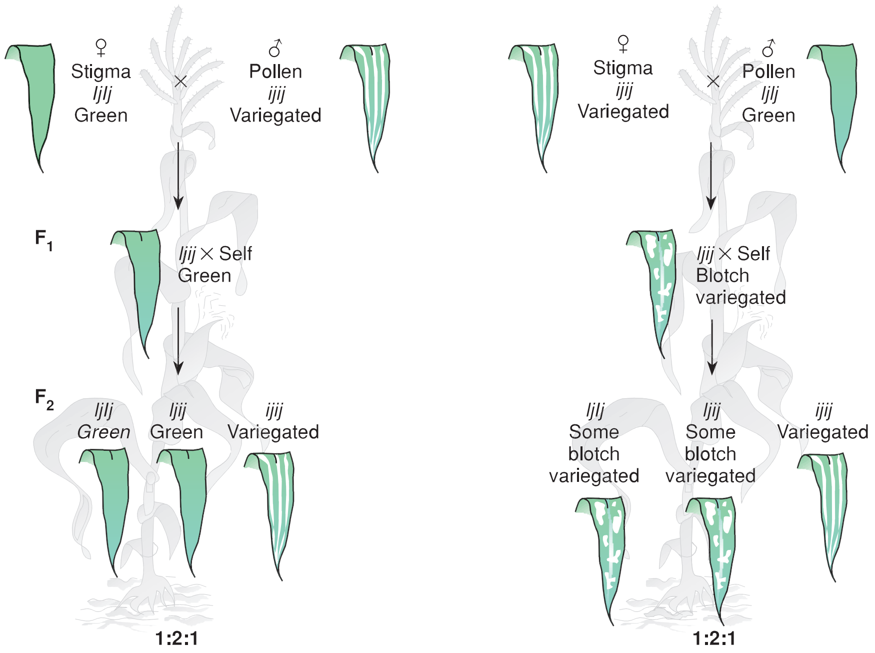


**Figure 17.9** Reciprocal crosses involving the chromosomal gene *iojap* in corn. The homozygous recessive condition (*ijij*) induces variegation (representative corn leaves are shown). (Blotch variegation consists of irregularly shaped *white* areas rather than striping.) However, plants with the dominant allele (*IjIj, Ijij*) can still be variegated if their mothers were variegated, since mothers pass on their chloroplasts to their offspring; males (pollen parents) do not pass on their chloroplasts. *Iojap* homozygotes induce variegation. The defective chloroplasts are then inherited in a cytoplasmic fashion.

plastids, they do not revert to normal type even under the *Ij* allele. Thus, we see the interaction of a chromosomal gene and the chloroplast itself, which "inherits" a changed condition.

There is some evidence that *iojap* may suppress the chloroplast rather than cause a mutation of some function. There are loci in corn and in other species that can induce back mutation in the chloroplasts. Removal of suppression rather than an actual reversion is more likely to occur because the reversion rate is too high to be due to simple back mutation.

### *Four-O'clocks*

The first work with corn variegation was done by Carl Correns, one of Mendel's rediscoverers. Correns also found maternal inheritance of variegation in the four-o'clock plant, *Mirabilis jalapa.* He could predict color and variegation of offspring solely on the basis of the region of the plant on which the stigma parent was located. A flower from a white sector, when pollinated by any pollen, would produce white plants; a flower on a green sector or a variegated sector produced green or variegated plants, respectively, when pollinated by pollen from any region of a plant. We thus see the simple maternal nature of the inheritance of the variegation. A chromosomal gene, like *iojap,* induces variegation. Inheritance of this induced variegation follows the "maternal" pattern of chloroplast inheritance.

## Chlamydomonas

The single-celled green alga, *Chlamydomonas reinhardi,* has been used in the study of extrachromosomal inheritance for several reasons. First, it has a single, large chloroplast; it can survive by culture technique even when the chloroplast is not functioning; and finally, it shows some interesting non-Mendelian inheritance patterns related to mating type. R. Sager has done extensive work on the inheritance of streptomycin resistance in *Chlamydomonas.*

Streptomycin resistance can be selected for in *Chlamydomonas* in several ways. Normal cells, sensitive to the antibiotic, are killed in its presence. If cells are grown in low levels of the antibiotic (100 g/ml), some cells show resistance to it. When these cells are crossed

Ruth Sager (1918–1997).
(Courtesy of Dr. Ruth Sager.)

with the wild-type, the resistance segregates in a 1:1 ratio, indicating that streptomycin resistance is controlled by a chromosomal locus. The same experiment can be repeated using high levels of the antibiotic in the medium (500–1,600 g/ml). Again, resistant colonies grow. If they are crossed with the wild-type, a 1:1 ratio does not ensue.

*Chlamydomonas* does not have sexes but does have mating types $mt^+$ and $mt^-$. Only individuals of opposite type can mate. Mating type is inherited as a single locus with two alleles. When two haploid cells of opposite mating type fuse, they form a diploid zygote, which then undergoes meiosis to produce four haploid cells, two of $mt^+$ and two of $mt^-$. The high-level resistance always segregates with the $mt^+$ parent (fig. 17.10). It is as if the $mt^+$ parent were contributing the cytoplasm to the zygote in a manner similar to maternal plastid inheritance in plants. The $mt^-$ parent acts like a pollen parent by making a chromosomal contribution but not a cytoplasmic one.

The mechanism of the extrachromosomal inheritance pattern of *Chlamydomonas* is the preferential digestion of the DNA of the chloroplast from the $mt^-$ parent. Currently, we believe that streptomycin's target is the chloroplast.

More recent work has shown that the $mt^+$ inheritance is only 99.98% effective—that is, 0.02% of the offspring in crosses of the type shown in figure 17.10 have the streptomycin phenotype of the $mt^-$ parent. Thus, we have the possibility of studying recombination in chloroplast genes. Although most of the evidence is only indirect and plagued by the previously mentioned problems of separating chloroplast and mitochondrial effects, some initial mapping studies have been done.

## Infective Particles

### Paramecium

Tracy Sonneborn discovered the killer trait in *Paramecium.* Before analyzing this trait, we must digress a moment to look at the life cycle of *Paramecium,* a ciliated

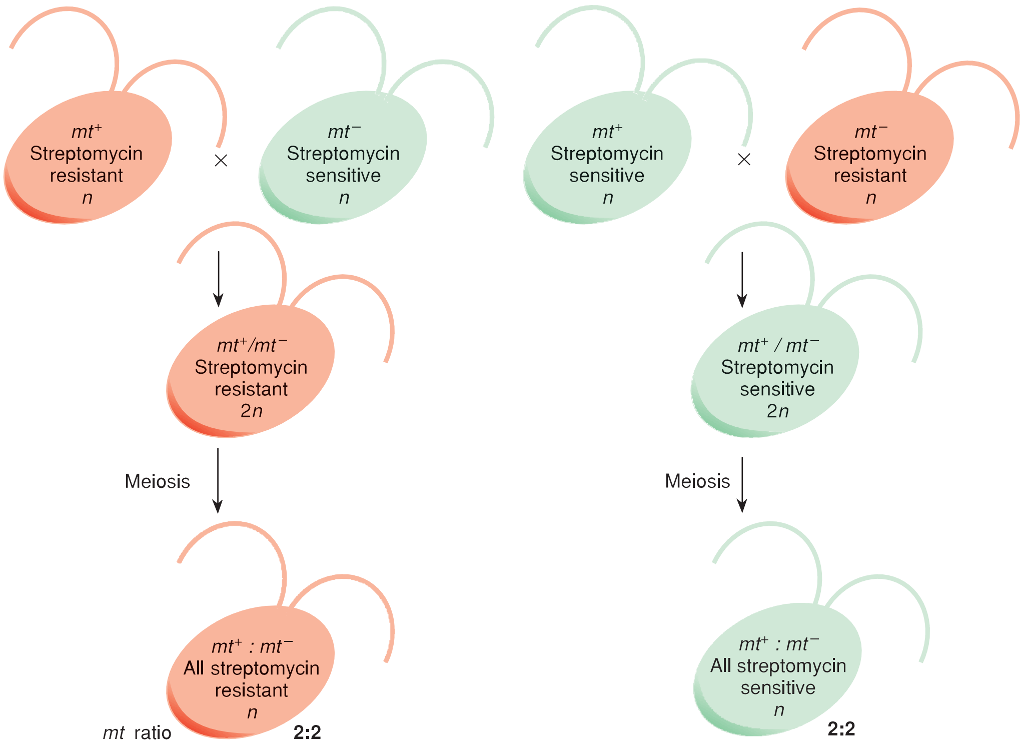

mt+
Streptomycin
resistant
n
×
mt−
Streptomycin
sensitive
n
mt+
Streptomycin
sensitive
n
×
mt−
Streptomycin
resistant
n
mt+/mt−
Streptomycin
resistant
2n
mt+ / mt−
Streptomycin
sensitive
2n
Meiosis
Meiosis
mt+ : mt−
All streptomycin
resistant
n
mt+ : mt−
All streptomycin
sensitive
n
mt ratio
2:2
2:2

**Figure 17.10** Inheritance pattern of streptomycin resistance in *Chlamydomonas* is dependent on the genotype of the $mt^+$ parent. (The $n$ and $2n$ refer to the ploidy of the cells.) If the $mt^+$ parent is streptomycin resistant *(red),* then the diploid heterozygote, as well as the meiotic products, will be streptomycin resistant. If, however, the $mt^+$ parent is streptomycin sensitive *(green),* the diploid heterozygote, as well as the meiotic products, will be streptomycin sensitive.

Tracy M. Sonneborn (1905–1981). (Courtesy: Indiana University Archives.)

protozoan familiar to most biologists. Ciliates have two types of nuclei: macronuclei and micronuclei. In *Paramecium,* there are two micronuclei, which are primarily reproductive nuclei, and one macronucleus, which is a polyploid nucleus concerned with the vegetative functions of the cell. During cell division, termed **binary fission,** the micronuclei divide by mitosis and the macronucleus constricts and is pulled in half.

*Paramecium* undergoes two types of nuclear rearrangements, during conjugation and **autogamy.** In conjugation, individuals of two mating types come together and form a connecting bridge. The nuclear events are shown in figure 17.11. Briefly, the macronucleus of each cell disintegrates while the micronuclei undergo meiosis. Of the resulting eight micronuclei per cell, seven disintegrate and one remains; this one undergoes mitosis to form two haploid nuclei per cell. A reciprocal exchange of nuclei across the bridge then occurs. Each cell now has two haploid nuclei, one original and one migrant. The two nuclei fuse to form a diploid nucleus. The diploid nuclei in the two conjugating cells are genetically identical because of the reciprocity of the process. These nuclei then undergo two mitoses each to form four diploid nuclei per cell. Two nuclei become macronuclei, which separate at the next cell division; two remain as micronuclei that divide by mitosis at the next cell division. The two cells that separate are known as **exconjugants.** Depending primarily on the amount of time conjugating cells remain united, an exchange of cytoplasm may occur along with the exchange of nuclei.

In the second type of process, autogamy, only one *Paramecium* is involved (fig. 17.12). The nuclear events are the same as in conjugation except that, at the point where a reciprocal exchange of nuclei would take place, the two haploid nuclei within the cell fuse. All cells after autogamy are homozygous.

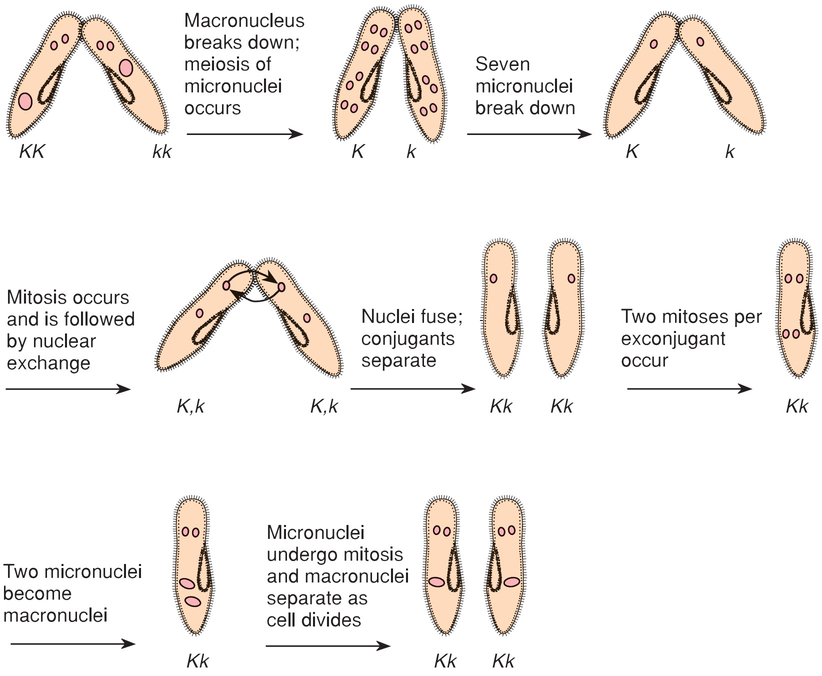

KK
kk
Macronucleus breaks down; meiosis of micronuclei occurs
K
k
Seven micronuclei break down
K
k
Mitosis occurs and is followed by nuclear exchange
K,k
K,k
Nuclei fuse; conjugants separate
Kk
Kk
Two mitoses per exconjugant occur
Kk
Two micronuclei become macronuclei
Kk
Micronuclei undergo mitosis and macronuclei separate as cell divides
Kk
Kk

**Figure 17.11** Conjugation in *Paramecium.* The letters *K* and *k* represent alleles of a gene in each micronucleus. When a *KK* and a *kk* individual conjugate, the exconjugants have the identical *Kk* genotype.

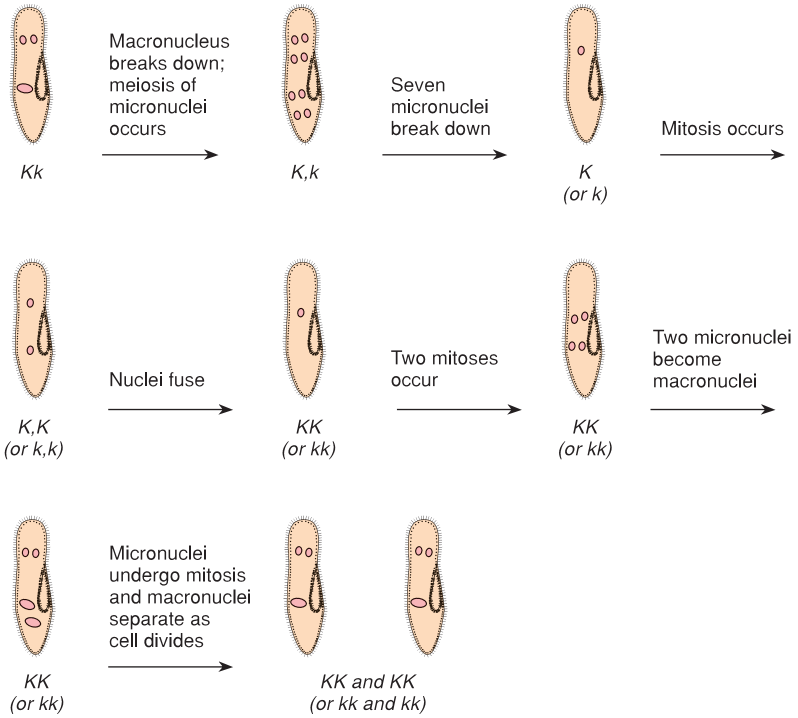

Kk
Macronucleus breaks down; meiosis of micronuclei occurs
K,k
Seven micronuclei break down
K (or k)
Mitosis occurs
K,K (or k,k)
Nuclei fuse
KK (or kk)
Two mitoses occur
KK (or kk)
Two micronuclei become macronuclei
KK (or kk)
Micronuclei undergo mitosis and macronuclei separate as cell divides
KK and KK (or kk and kk)

**Figure 17.12** Autogamy in *Paramecium.* The letters *K* and *k* represent alleles of a gene in each micronucleus. If a heterozygote undergoes autogamy, it becomes homozygous for one of the alleles (*KK* or *kk*).

### *Killer* Paramecium *and Kappa Particles*

Sonneborn and his colleagues found that when certain stocks of *Paramecium* were mixed together, one stock had the ability to cause the individuals of the other stock to die. Those individuals causing death were called "killers" and those dying were referred to as "sensitives." During conjugation, the sensitives are temporarily resistant to the killers. If cytoplasm is not exchanged during the conjugation, the exconjugants retain their original phenotypes so that killers stay killers and sensitives stay sensitives. When an exchange of cytoplasm occurs between sensitive and killer cells, both exconjugants are killers. The transfer of some cytoplasmic particle seems to be implied. Indeed, Sonneborn observed such particles in the cytoplasm of killers and called them **kappa particles** (fig. 17.13).

Although the occurrence of killer *Paramecium* does not appear to involve chromosomal genes, Sonneborn reported one case in which exconjugant killer paramecia of hybrid origin underwent autogamy. He found that half of the resulting cells had no kappa particles and had become sensitives. He concluded that a gene is required for the presence of kappa particles, which has subsequently been verified by numerous crosses. Figure 17.14 illustrates the sequence of genetic events that would produce a heterozygous killer *Paramecium* that, upon autogamy, would have a 50% chance of becoming sensitive.

Although not yet cultured outside of a *Paramecium,* kappa is presumably a bacterium because it has many bacterial attributes including size, cell wall, presence of DNA, and presence of certain prokaryotic reactions (fig. 17.15). J. Preer and his colleagues, who studied kappa itself, named it *Caedobacter taeniospiralis.* Kappa occurs in at least two forms. The *N* form, the infective form that passes from one *Paramecium* to another, does not confer killer specificity on the host cell. The *N* form is attacked by bacteriophages that induce formation of inclusions, called *R* bodies, inside the kappa particle and thus convert it to the *B* form. These *R* bodies are visible under the light microscope as refractile bodies (fig. 17.15).

In the *B* form, kappa can no longer replicate; it is often lysed within the cell. It confers killer specificity on the host cell, however. The sensitives are killed by the toxin **paramecin,** which is released by the killer *Paramecium* into the environment. Precisely what steps are involved in its formation are not known, although it is

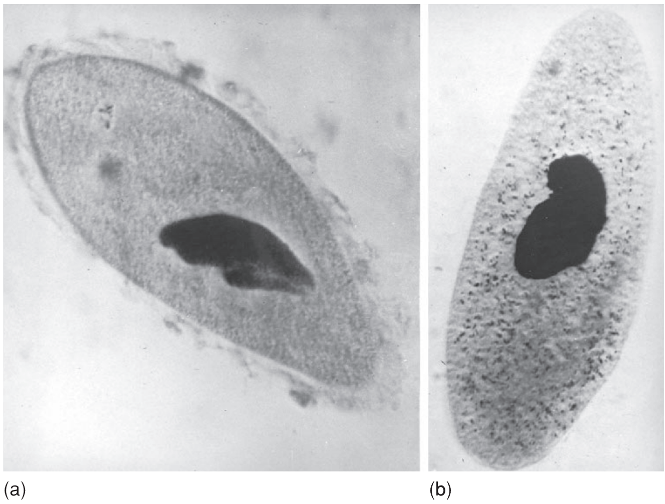


**Figure 17.13** (*a*) Normal (sensitive) *Paramecium.* (*b*) Kappa-containing (killer) *Paramecium.* A *Paramecium* is about 200 μm long. (Source: T. M. Sonneborn, figure 29.3, p. 373 in I. H. Herskowitz, *Genetics,* 2nd ed. [Boston: Little, Brown, 1965]. Reproduced by permission.)

plain that the virus plays an integral role. Whether the viral DNA or the kappa DNA codes the toxin is also not known at present.

### Mate-Killer Infection and Mu Particles

Kappa is not the only infective agent known in *Paramecium.* Another agent is the **mate-killer** infection. Here again, killer cells have visible, bacterialike particles, called **mu particles,** in the cytoplasm. Preer and his colleagues have named them *Caedobacter conjugatus.* Mate-killers do not release a toxin into the environment, but instead kill their mates during conjugation. One of two unlinked dominant genes, $M_1$ and $M_2$, is required for the presence of mu particles. An interesting phenomenon occurs when a mate-killer becomes homozygous $m_1m_1\ m_2m_2$ by autogamy. Although the offspring eventually lose their mu particles, virtually no loss of particles occurs until about the eighth generation, when some offspring lose all their mu. Up to this generation, all the cells maintain a full complement of mu. In the fifteenth generation, only about 7% of the cells still have mu particles.

This phenomenon is explained as the diluting out not of the mu themselves, but of a factor called **metagon,** which is necessary for the maintenance of mu in the cell. Once the cell becomes homozygous recessive, no further metagon production occurs. The verification that metagon is subsequently diluted out is evident in fifteenth-generation cells that still have their mu. We would expect that after fission, one daughter cell would have a metagon and the other would not. What we expect, in fact, happens. The rate of dilution is consistent with an original number of about one thousand metagons per cell. The metagon appears to be

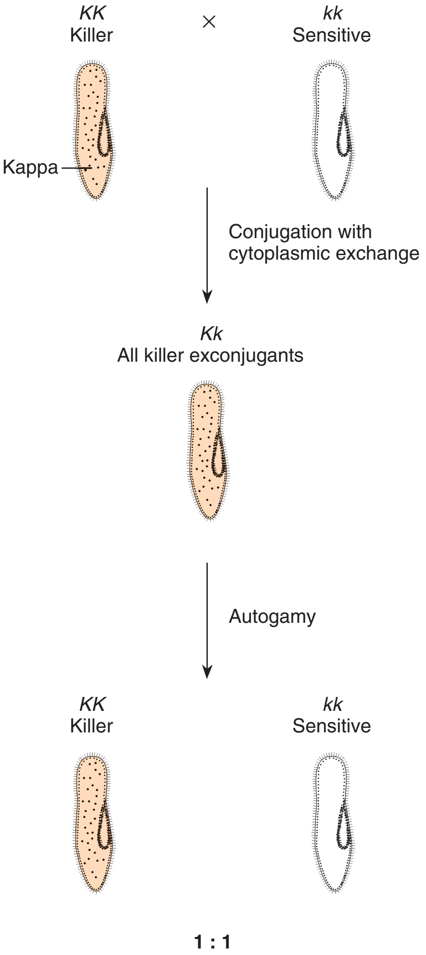


**Figure 17.14** Autogamy in a heterozygous (*Kk*) killer *Paramecium* (formed by conjugation, with cytoplasmic exchange, of a *KK* killer and a *kk* sensitive cell). Upon autogamy, the heterozygote has a 50% chance of becoming a homozygous (*KK*) killer or a homozygous (*kk*) sensitive cell that loses its kappa particles.

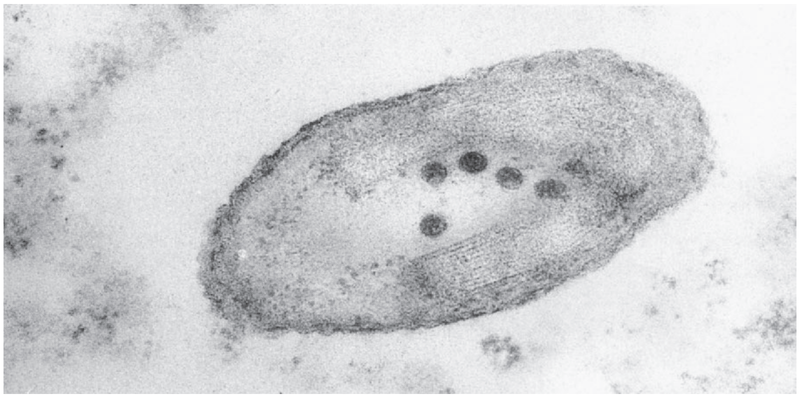


**Figure 17.15** Electron micrograph of a sectioned kappa particle (*Caedobacter taeniospiralis*). Phage particles appear as dark inclusions. The plane of the section cuts through a rolled-up *R* body. Magnification 61,200×. (Reproduced by permission of J. R. Preer, Jr.)

messenger RNA because it is destroyed by RNase. Its protein product is presently unknown.

We thus see several instances of infective particles that interact with the *Paramecium* genome to produce interesting phenotypic results. Similar interactions are known in other organisms—for example, the killer trait in yeast.

### Drosophila

Several infective particles mimic patterns of inheritance in insects. In *Drosophila,* we find forms of the **sex-ratio phenotype** in which females produce mostly, if not exclusively, daughters. One form is inherited as a chromosomal gene; another form, however, is not chromosomal. In the nonchromosomal form, females usually produce a few sons. These sons do not pass on the sex-ratio trait, but the daughters of sex-ratio females do. Because the trait persisted even after all the chromosomes had been substituted out of the stock by appropriate crosses, it was proven to be extrachromosomal.

In addition, about half the eggs of a sex-ratio female fail to develop. Cytoplasm can be withdrawn from the undeveloped eggs and used to infect other females. The trait, then, is caused by some cytoplasmic factor that could infect other females and is not passed on by sperm. Detailed cytological examination of the cytoplasm of sex-ratio females has revealed a spirochete (fig. 17.16) that has been isolated and used to infect other female *Drosophila* with the sex-ratio trait; it is, therefore, the causal agent of this phenotype.

## Prokaryotic Plasmids

In chapters 7 and 13, we discussed the role of plasmids in the study of prokaryotic genetics and in recombinant DNA work. They are mentioned again here because they represent extrachromosomal genetic systems, primarily in prokaryotes. The autonomous segments of DNA known as plasmids are, for the most part, known from bacteria, in which they occur as circles of DNA within the host cell (noncircular DNA is soon degraded). When plasmids become integrated into the chromosomes, they become indistinguishable from chromosomal material.

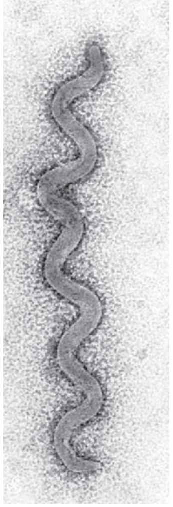


**Figure 17.16** Electron micrograph of the spirochete associated with the extrachromosomal sex-ratio trait in *Drosophila.* Magnification 22,700×. (K. Oishi and D. F. Poulson, "A virus associated with SR-spirochetes of *Drosophila nebulosa,*" *Proceedings of The National Academy of Sciences, USA,* 67 [1970]:1565–72. Reproduced by permission of the authors.)

### R *and* Col Plasmids

In addition to the F factor found frequently in bacteria, a variety of other plasmids occur, including the R and Col plasmids. The **R plasmids** carry genes for resistance to various antibiotics, and the **Col plasmids** have genes that are responsible for producing proteins called *colicins,* which are toxic to strains of *E. coli* (fig. 17.17). Plasmids containing genes for Col-like toxins specific for other bacterial species are also known. Col and R plasmids can exist in two states. In one state, the plasmid has

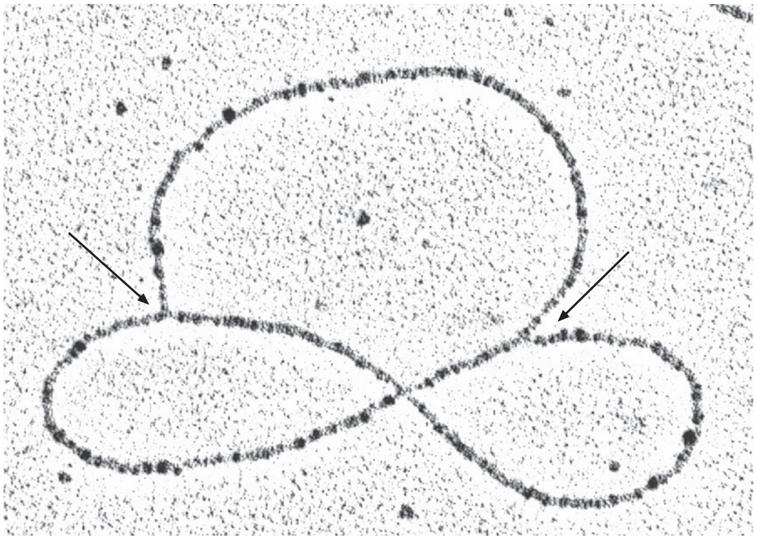


**Figure 17.17** Electron micrograph of replication of Col E1 circular plasmid. The *arrows* mark the branch points of the theta structure. Magnification 90,000×. (Source: J. I. Tomizawa, Y. Sakakibara, and T. Kakefuda, "Replication of Colicin E1 plasmid DNA in cell extracts: Origin and direction of replication," *Proceedings of The National Academy of Sciences, USA,* 71 [1974]:2260–64.)

a sequence of genes called the **transfer operon *(tra)*,** which makes the plasmids similar to F factors in that they can transfer their genes from one bacterium to the next. In the other state, the plasmids lack this operon and cannot transfer their loci to another cell. Thus, Col and R plasmids are actually made of two parts: the loci for antibiotic resistance or colicin production and the part responsible for infectious transfer. In R plasmids, the infectious transfer part is called the **resistance transfer factor** (RTF).

The occurrence of resistance plasmids was observed in Japan in the late 1950s, when it was discovered that bacteria were simultaneously acquiring resistance to several antibacterial agents. When cultures of *Shigella,* a dysentery-causing bacterium, were exposed to streptomycin, sulfonamide, chloramphenicol, or tetracycline, the bacteria exhibited resistance not only to the one particular agent they were exposed to but to one or more of the others as well. The plasmid responsible for this multiple resistance was named R222.

The Col plasmids contain loci that produce proteins that are toxic, for various reasons, to strains of bacteria not carrying the plasmids. Colicins attack sensitive bacterial cells at bacterial surface receptors. They have been classified into twenty or more categories according to the types of receptors they attack. Some colicins may enter the cell directly, but others do not. For example, colicin K appears to kill sensitive cells by inhibiting DNA, RNA, and protein synthesis, although not directly entering the cell. Colicin E3, however, acts as an intracellular ribonuclease that cleaves off about fifty nucleotides from the 3′ end of the 16S ribosomal RNA within the ribosome. The cleavage inactivates the sensitive cell's ribosomes and is, of course, lethal.

Since many R plasmids, Col plasmids, and F factors, as well as host chromosomes, have insertion sequences (chapter 14), a good deal of exchange occurs among the plasmids, and many are able to integrate into the host chromosome. Although their mobility makes it easier to map and study plasmids, it also poses a human health problem. Resistance to various antibacterial agents is easily transferred among enterobacteria worldwide. This can even occur outside of host organisms (people) where pollution or sewage is found. In addition, resistance found in relatively harmless enterobacteria, such as *E. coli,* can easily pass to more pathogenic bacteria, such as *Shigella* and *Salmonella.* Since we are selecting for resistance every time we use antibacterial drugs, we should not use these drugs indiscriminately. For some time, health workers have been concerned about excessive medical use of antibacterial drugs as well as about the large quantities of antibiotics used in animal feed.

### Uncovering Plasmids

How do we know when the phenotype is controlled by a plasmid rather than by the chromosomal genes of a bacterium? Plasmids can be seen with an electron microscope or by density-gradient centrifugation of the cell's DNA. But several less direct lines of evidence also supply the answer. To begin with, multiple aspects of the phenotype (e.g., resistance to several antibacterial agents) change simultaneously, as with plasmid R222. Another clue is that the phenotypic change is infectious: Japanese workers found that with R222, resistant cells converted nonresistant cells. As B. Lewin stated, "Resistance is infectious."

Several other clues point to the presence of a plasmid. In linkage studies, using transduction for example, plasmid loci show no linkage to host loci; plasmids themselves can be mapped because their loci are linked to each other. Since the plasmid DNA replicates at its own speed, it can miss being incorporated into a daughter cell. Thus, many spontaneous losses of the plasmid occur. And finally, certain treatments—with acridine dyes, for example—have little effect on the replication of the host chromosome, but selectively prevent the plasmid from replicating; thus the plasmid can be eliminated from the cell population. The existence of plasmids in a bacterial population can, therefore, be verified with morphological, physiological, and analytical evidence.

## IMPRINTING

Although sex linkage alters inheritance patterns, we do not expect different inheritance patterns, dependent on the parent of origin, from genes located on autosomal chromosomes. That is, the genotype of an offspring should be predicted by its alleles regardless of which parent donated which allele. That understanding has now been shown to be incorrect for a group of genes whose phenotypic effects are determined by the parent that donated a particular allele. This phenomenon is called *imprinting* (or *molecular imprinting* or *parental imprinting*). It falls under the general classification of an **epigenetic effect**, a term that has come to mean an effect due to an environmentally induced change in the genetic material but not causing a change in base pairs. It is a phenomenon of differential expression of the alleles at a locus depending on which parent the gene originated with.

A striking example of imprinting in human beings involves two medical syndromes, both resulting in mental retardation. In Prader-Willi syndrome, affected persons are extremely obese; in Angelman syndrome, those affected are thin and sometimes referred to as "happy puppets," because they exhibit a happy facial expression and erratic, jerky movements. It turns out that both syndromes are associated with deletions in the long arm of chromosome 15, in bands 15q11–q13. The effect is seen in an individual arising from a gamete missing a 15q11–q13 region. If the remaining region is of paternal origin, due to a deletion of the maternal gene, the offspring will have Angelman syndrome; if the remaining region is of maternal origin, the offspring will have Prader-Willi syndrome. This unusual situation indicates that the phenotype is dependent on the parent from which the region comes. Recently, this region of chromosome 15 has come under intense scrutiny. In males, from five to seven genes are expressed from this area, and in females, one gene has been identified as the cause of Angelman syndrome, *UBE3A* (E3 ubiquitin protein ligase). It has been hypothesized that there is an **imprinting center (IC),** a region responsible for the control of imprinting. The imprinting mark is almost certainly DNA methylation, which has the property of turning off gene transcription. Stretches of CG repeats (called **CpG islands,** in which CpG indicates sequential bases on the same strand of DNA rather than a C-G base pair) have been found in these imprinting centers. The imprinting center would be the site of the erasure of past imprinting and the initiation of new imprinting during gametogenesis. Over twenty genes exhibit imprinting, and the epigenetic phenomenon also appears in proteins, with differential acetylation of proteins as the imprinting mark.

The question arises as to how imprinting evolved; that is, what evolutionary advantages come from silencing an allele from one of the parents? Although we don't really know at this point, several hypotheses have been suggested, including competition among maternal and paternal alleles for expression (see chapter 21). For example, the *Igf-2* gene (insulin-like growth factor) places demands on pregnant females to produce larger fetuses. This is advantageous to the father (assuming that the female will have offspring from several fathers), but not the mother. So, the mother's gene is usually methylated and therefore inactive. It is as if the genes are in competition with each other, with the father's gene promoting the formation of a large fetus and the mother's gene promoting the formation of a smaller fetus. Currently, the phenomenon of imprinting is under active study.

## SUMMARY

**STUDY OBJECTIVE 1:** To analyze the inheritance patterns of maternal effects 509–510

Patterns of non-Mendelian inheritance fall into two categories: maternal effects and cytoplasmic inheritance. Maternal effects are illustrated by snail-shell coiling. The direction of coiling is determined by the genotype of the maternal parent, with dextral coiling dominant to sinistral coiling.

**STUDY OBJECTIVE 2:** To analyze the patterns of cytoplasmic inheritance 511–524

Cytoplasmic inheritance is usually seen in organelles, symbionts, or parasites that have their own genetic material. Chloroplasts and mitochondria have relatively small, circular chromosomes with prokaryotic affinities. An interaction exists between organelles and nuclei; the organelles do not encode all their own proteins and enzymes. Mitochondrial defects can be inherited through nuclear genes or through the mitochondrion itself. A similar pattern is seen in chloroplasts. The processes of cytoplasmic inheritance are exemplified by symbiotic bacteria in *Paramecium.*

Plasmids are autonomous segments of DNA. In prokaryotes, R and Col plasmids, as well as the F factor, have been well studied. Plasmids usually carry an operon for transfer and insertion sequences for attachment to cell chromosomes and to each other. Hence, they represent highly mobile segments of genetic material.

**STUDY OBJECTIVE 3:** To analyze the patterns of imprinting 524

Imprinting is a phenomenon of gene activity affected by the parent of origin. Due to a pattern of gene methylation that differs in male and female parents, a gene may show differential activity depending on the parent from which it came. More than twenty genes exhibiting this epigenetic phenomenon are known.

## SOLVED PROBLEMS

**PROBLEM 1:** What possible phenotypes and genotypes could the female parent of a sinistrally coiled snail have?

***Answer:*** If a snail is sinistrally coiled, its mother must have had the *dd* genotype, since sinistrality is recessive. If the female parent is a recessive homozygote, its mother must have contributed a recessive *d* allele. Therefore *its* mother (the grandmother) could have had either a *Dd* or *dd* genotype. Its daughter could therefore be either dextrally or sinistrally coiled (respectively). Thus, to answer the question, a sinistrally coiled snail could have had a mother that was either dextrally or sinistrally coiled, but only of the *dd* genotype.

**PROBLEM 2:** You have just noticed a petite yeast colony growing in a petri plate under aerobic conditions. What type of petite is it?

***Answer:*** The simplest way to determine the nature of the lesion resulting in the petite phenotype is to make a cross of the petite strain with a wild-type strain. After meiosis, isolate the four products (spores) and allow them to grow separately under normal, aerobic conditions. If the ratio of petite to wild-type is 1:1, the mutation is of a nuclear gene. If progeny are wild-type, the mutation is in the mitochondrial genome and is of the neutral type. If progeny are mostly petites, the mutation is also in the mitochondrial genome, but it is of the suppressive type.

**PROBLEM 3:** Killer *Paramecium* with the genotype *KK* are mated with *kk* cells under a situation that allows cytoplasmic exchange. If the exconjugants undergo autogamy, what types of progeny would you expect?

***Answer:*** Both exconjugants will be *Kk,* and since cytoplasmic exchange occurred, both cytoplasms will contain kappa. Autogamy will produce either *KK* or *kk* cells. Since at least one *K* gene is needed for the maintenance of kappa, the *kk* cells eventually lose the kappas and become sensitive. Thus, we expect 1/2 sensitive:1/2 killers.

## EXERCISES AND PROBLEMS*

**DETERMINING NON-MENDELIAN INHERITANCE**

1. J. Christian and C. Lemunyan have shown that mice raised under crowded conditions produce two generations with reduced growth rates. What sort of genetic control might exist, and how could this control be demonstrated?
2. Describe the types of evidence that could be gathered to determine whether a trait in *E. coli* is controlled by chromosomal or plasmid genes. (*See also* CYTOPLASMIC INHERITANCE)
3. The maroon-like (*ma-l*) locus in *Drosophila* is inherited in an X-linked recessive fashion. If you cross a heterozygous female with a maroon-like male, all the progeny are wild-type. If the female progeny from this cross are mated again with maroon-like males, half of the females produce all maroon-like progeny, and the other half produce all wild-type progeny. Explain these results. (*See also* MATERNAL EFFECTS)

---

*Answers to selected exercises and problems are on page A-20.

### MATERIAL EFFECTS

4. Snail coiling is called a maternal trait. Is it possible that it is caused by an allele at a sex-linked locus?
5. How would you rule out a viral origin for snail-shell coiling?
6. Give the genotypes involved when a sinistral female snail produces dextral offspring. What genotypes could the male parent of the sinistral female have?
7. A dextral snail is self-fertilized and produces only sinistral progeny. What is the probable genotype of this snail and its parents?
8. In corn, male sterility is controlled by a maternal cytoplasmic element. A dominant nuclear gene, Restorer (*Rf*), restores fertility to male sterile lines. If pollen from a homozygous *RfRf* plant is used to pollinate a male sterile plant, what genotypes and phenotypes would you expect in the progeny?

### CYTOPLASMIC INHERITANCE

9. What evidence indicates that it is not absolutely essential, in an evolutionary sense, for mitochondria to have genes for specific components of oxidative phosphorylation?
10. How would you determine that a segregative petite mutant in yeast is controlled by a chromosomal gene?
11. What results would you obtain by making all possible pairwise crosses of the three types of yeast petites?
12. An ornamental spider plant has green and white striped leaves. How can you determine whether cytoplasmic inheritance is responsible for the striping and whether there is interaction with an *iojap*-type chromosomal gene?
13. In *Chlamydomonas,* 0.02% of the meiotic products are of the $mt^-$ parental type. How can you use this information in mapping? (Use streptomycin sensitivity, $str^s$, and resistance, $str^r$, as an example.)
14. What similarities do mitochondria and plastids share?
15. What evidence is there that mitochondria and chloroplasts originated from prokaryotes?
16. Individuals from killer and nonkiller strains of *Paramecium* are mixed together. Cytoplasmic exchange occurs during conjugation. Approximately 25% of the exconjugants are sensitive, and the remaining 75% are killers. What are the genotypes of the individuals of the two strains, and what ratios of sensitives and killers would result if the various exconjugants underwent autogamy?
17. What genetic tests could you conduct to show that the mate-killer phenotype in *Paramecium* requires a dominant allele at any one of *two* loci?
18. Resistant and sensitive strains of *Drosophila melanogaster* differ in their ability to tolerate $CO_2$—anesthetization with it kills sensitive flies. What genetic experiments would you perform to determine whether the trait is caused by a virus? How would you rule out chromosomal genes?
19. Suppose you have identified a person who has introns in his or her mitochondrial DNA. What would you deduce about the origin of this DNA?
20. A mutation in the mitochondrial genome in people causes blindness. If reciprocal matings between affected and normal individuals occur in a family pedigree, what types of children would you expect from each cross?
21. When chloroplast DNA from *Chlamydomonas* is digested with a particular restriction enzyme and then hybridized with a particular probe, two bands are detected. Some strains (type 1) yield bands of 1.5 and 3.7 kilobases; other strains (type 2) yield bands of 2.5 and 6.0 kilobases. For the following crosses, predict the progeny:
    **a.** $mt^+$, strain 1 × $mt^-$, strain 2
    **b.** $mt^+$, strain 2 × $mt^-$, strain 1
22. What type of asci do you expect if you cross a yeast strain carrying an antibiotic resistance gene in its mitochondria with a strain that has normal (sensitive) mitochondria?
23. In *Paramecium,* the maintenance of kappa particles requires the dominant nuclear gene *K.* A *Kk* killer cell conjugates with a sensitive cell of the same genotype without cytoplasmic exchange. Predict the genotypes and phenotypes that result if each exconjugant then undergoes autogamy.
24. In *Neurospora,* the slow-growing trait *poky* is inherited maternally and is due to an abnormal respiratory protein. A nuclear gene *F* makes *poky* individuals grow faster, even though the protein is still defective. Such strains are called *fast-poky* (*F′* is normal *poky*). *Poky* cytoplasm is not altered by *F* in a zygote, and *F* has no effect on normal cytoplasm. What genotypes and phenotypes do you expect if the maternal parent is *fast-poky* and the paternal parent is normal?
25. In corn, two independent, recessive nuclear genes, *japonica (j)* and *iojap (ij),* produce variegation (green and white striped leaves). Matings between individuals heterozygous for *japonica* always produce 3 green:1 striped individuals regardless of how the cross is performed. The behavior of *iojap* was described in figure 17.9. You have a variegated plant that could be either *jj* or *ijij.* What cross can you make to determine the genotype of this plant, and what results do you expect in the $F_1$ generation in each case?

**26.** If *Paramecium* cells heterozygous for both genes involved in the maintenance of the mate-killer trait are forced to undergo autogamy, what phenotypic ratios do you expect?

**27.** A petite yeast strain is crossed with a wild-type strain. What phenotypic ratio do you expect after meiosis if the petite is
**a.** nuclear?
**b.** suppressive?
**c.** neutral?

## CRITICAL THINKING QUESTIONS

**1.** When a eukaryotic cell divides, cell organelles such as mitochondria and chloroplasts are distributed to the daughter cells. What mechanisms might exist to ensure an even distribution of these organelles?

**2.** Lamarckian inheritance, the inheritance of acquired characteristics, is generally discounted as a major evolutionary mechanism (chapter 21). (For example, Lamarck suggested that the long neck of the giraffe came about by as giraffes stretched for food, followed by the inheritance of this longer, stretched neck.) Is the progression of the lysogenic state of *E. coli* from one generation to the next an example of Lamarckian inheritance? Why or why not?

*Suggested Readings for chapter 17 are on page B-18.*

# 18

# QUANTITATIVE INHERITANCE

## STUDY OBJECTIVES

1. To understand the patterns of inheritance of phenotypic traits controlled by many loci 531
2. To investigate the way that geneticists and statisticians describe and analyze normal distributions of phenotypes 535
3. To define and measure heritability, the unit of inheritance of variation in traits controlled by many loci 542

## STUDY OUTLINE

**Traits Controlled by Many Loci** 531
  Two-Locus Control 531
  Three-Locus Control 532
  Multilocus Control 532
  Location of Polygenes 533
  Significance of Polygenic Inheritance 534
**Population Statistics** 535
  Mean, Variance, and Standard Deviation 536
  Covariance, Correlation, and Regression 539
**Polygenic Inheritance in Beans** 541
**Selection Experiments** 541
**Heritability** 542
  Realized Heritability 542
  Partitioning of the Variance 543
  Measurement of Heritability 544
**Quantitative Inheritance in Human Beings** 545
  Skin Color 545
  IQ and Other Traits 546
*Summary* 547
*Solved Problems* 548
*Exercises and Problems* 549
*Critical Thinking Questions* 551
**Box 18.1 Mapping Quantitative Trait Loci** 537
**Box 18.2 Human Behavioral Genetics** 547

Much human variation is quantitative.

(© Jim Cummins/FPG International.)

When we talked previously of genetic traits, we were usually discussing traits in which variation is controlled by single genes whose inheritance patterns led to simple ratios. However, many traits, including some of economic importance—such as yields of milk, corn, and beef—exhibit what is called *continuous variation.*

Although some variation occurred in height in Mendel's pea plants, all of them could be scored as either tall or dwarf; there was no overlap. Using the same methods that Mendel used, we can look at ear length in corn (fig. 18.1). With Mendel's peas, all of the $F_1$ were tall. In a cross between corn plants with long and short ears, all of the $F_1$ plants have ears intermediate in length between the parents. When both pea and corn $F_1$ plants are self-fertilized, the results are again different. In the $F_2$ generation, Mendel obtained exactly the same height categories (tall and dwarf) as in the parental generation. Only the ratio was different—3:1.

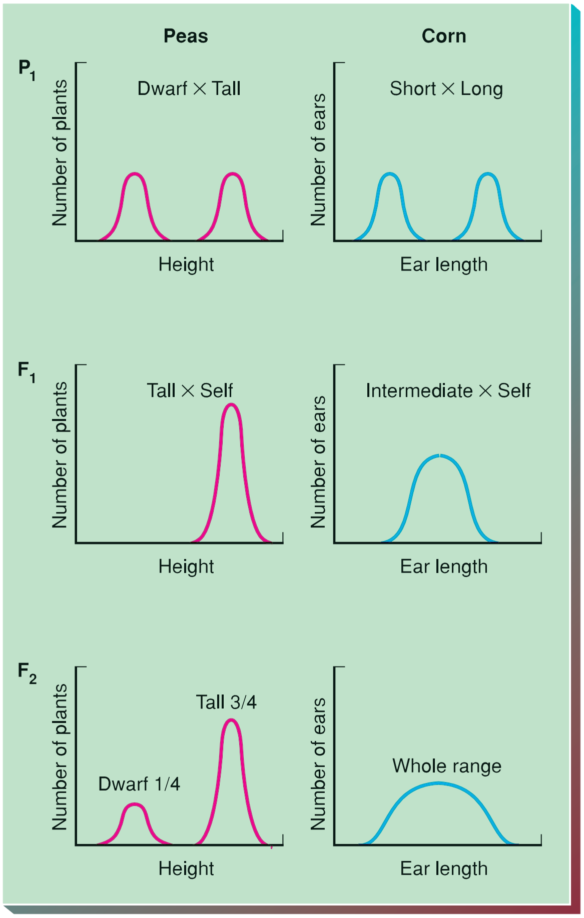


**Figure 18.1** Comparison of continuous variation (ear length in corn) with discontinuous variation (height in peas).

In corn, however, ears of every length, from the shortest to the longest, are found in the $F_2$; there are no discrete categories. A genetically controlled trait exhibiting this type of variation is usually controlled by many loci. In this chapter, we study this type of variation by looking at traits controlled by progressively more loci. We then turn to the concept of heritability, which is used as a statistical tool to evaluate the genetic control of traits determined by many loci.

## TRAITS CONTROLLED BY MANY LOCI

Let us begin by considering grain color in wheat. When a particular strain of wheat having red grain is crossed with another strain having white grain, all the $F_1$ plants have kernels intermediate in color. When these plants are self-fertilized, the ratio of kernels in the $F_2$ is 1 red:2 intermediate:1 white (fig. 18.2). This is inheritance involving one locus with two alleles. The white allele, $a$, produces no pigment (which results in the background color, white); the red allele, $A$, produces red pigment. The $F_1$ heterozygote, $Aa$, is intermediate (incomplete dominance). When this monohybrid is self-fertilized, the typical 1:2:1 ratio results. (For simplicity, we use dominant-recessive allele designations, $A$ and $a$. Keep in mind, however, that the heterozygote is intermediate in color.)

### Two-Locus Control

Now let us examine the same kind of cross using two other stocks of wheat with red and white kernels. Here, when the resulting intermediate (medium-red) $F_1$ are self-fertilized, five color classes of kernels emerge in a ratio of 1 dark red:4 medium dark red:6 medium red:4 light red:1 white (fig. 18.3). The offspring ratio, in sixteenths, comes from the self-fertilization of a dihybrid in which the two loci are unlinked. In this case, both loci affect the same trait in the same way. In figure 18.3, each capital

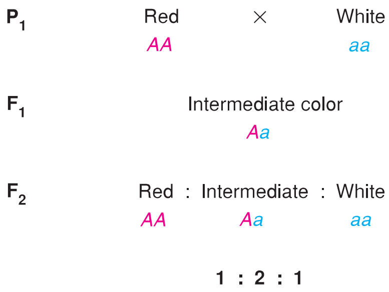


**Figure 18.2** Cross involving the grain color of wheat in which one locus is segregating.

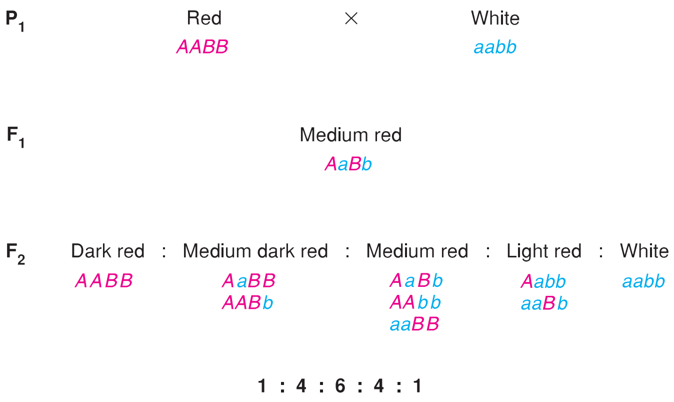
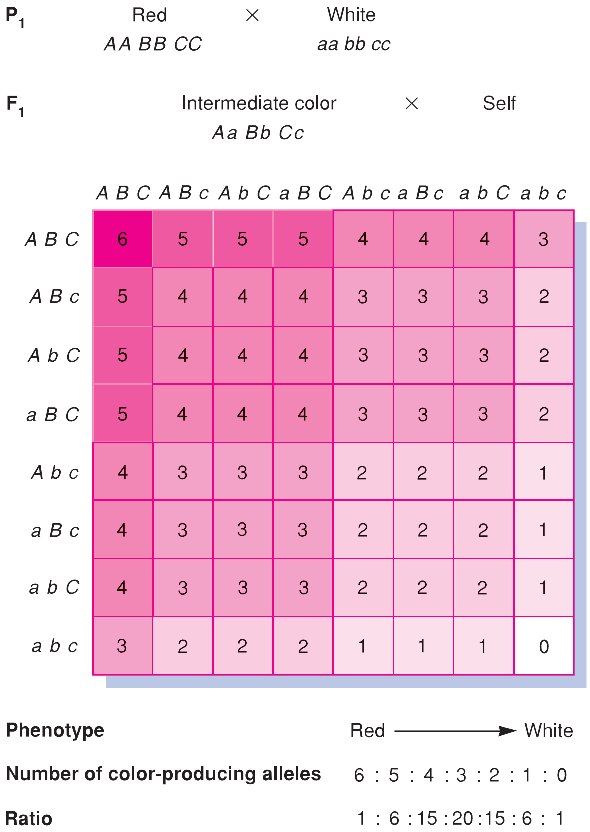


**Figure 18.3** Another cross involving wheat grain color in which two loci are segregating.

letter represents an allele that produces one unit of color, and each lowercase letter represents an allele that produces no color. Thus, the genotype *AaBb* has two units of color, as do the genotypes *AAbb* and *aaBB*. All produce the same intermediate grain color. Recall from chapter 2 that a cross such as this produces nine genotypes in a ratio of 1:2:1:2:4:2:1:2:1. If these classes are grouped according to numbers of color-producing alleles, as shown in figure 18.3, the 1:4:6:4:1 ratio appears. This ratio is a product of a binomial expansion.

## Three-Locus Control

In yet another cross of this nature, H. Nilsson-Ehle in 1909 crossed two wheat strains, one with red and the other with white grain, that yielded plants in the $F_1$ generation with grain of intermediate color. When these plants were self-fertilized, at least seven color classes, from red to white, were distinguishable in a ratio of 1:6:15:20:15:6:1 (fig. 18.4). This result is explained by assuming that three loci are assorting independently, each with two alleles, so that one allele produces a unit of red color and the other allele does not. We then see seven color classes, from red to white, in the 1:6:15:20:15:6:1 ratio. This ratio is in sixty-fourths, directly from the 8 × 8 (trihybrid) Punnett square, and comes from grouping genotypes in accordance with the number of color-producing alleles they contain. Again, the ratio is one that is generated in a binomial distribution.



| | *A B C* | *A B c* | *A b C* | *a B C* | *A b c* | *a B c* | *a b C* | *a b c* |
|---|---|---|---|---|---|---|---|---|
| *A B C* | 6 | 5 | 5 | 5 | 4 | 4 | 4 | 3 |
| *A B c* | 5 | 4 | 4 | 4 | 3 | 3 | 3 | 2 |
| *A b C* | 5 | 4 | 4 | 4 | 3 | 3 | 3 | 2 |
| *a B C* | 5 | 4 | 4 | 4 | 3 | 3 | 3 | 2 |
| *A b c* | 4 | 3 | 3 | 3 | 2 | 2 | 2 | 1 |
| *a B c* | 4 | 3 | 3 | 3 | 2 | 2 | 2 | 1 |
| *a b C* | 4 | 3 | 3 | 3 | 2 | 2 | 2 | 1 |
| *a b c* | 3 | 2 | 2 | 2 | 1 | 1 | 1 | 0 |

| | |
|---|---|
| **Phenotype** | Red ⟶ White |
| **Number of color-producing alleles** | 6 : 5 : 4 : 3 : 2 : 1 : 0 |
| **Ratio** | 1 : 6 : 15 : 20 : 15 : 6 : 1 |

**Figure 18.4** One of Nilsson-Ehle's crosses involving three loci controlling wheat grain color. Within the Punnett square, only the number of color-producing alleles is shown in each box to emphasize color production.

## Multilocus Control

From here, we need not go on to an example with four loci, then five, and so on. We have enough information to draw generalities. It should not be hard to see how discrete loci can generate a continuous distribution (fig. 18.5). Theoretically, it should be possible to distinguish

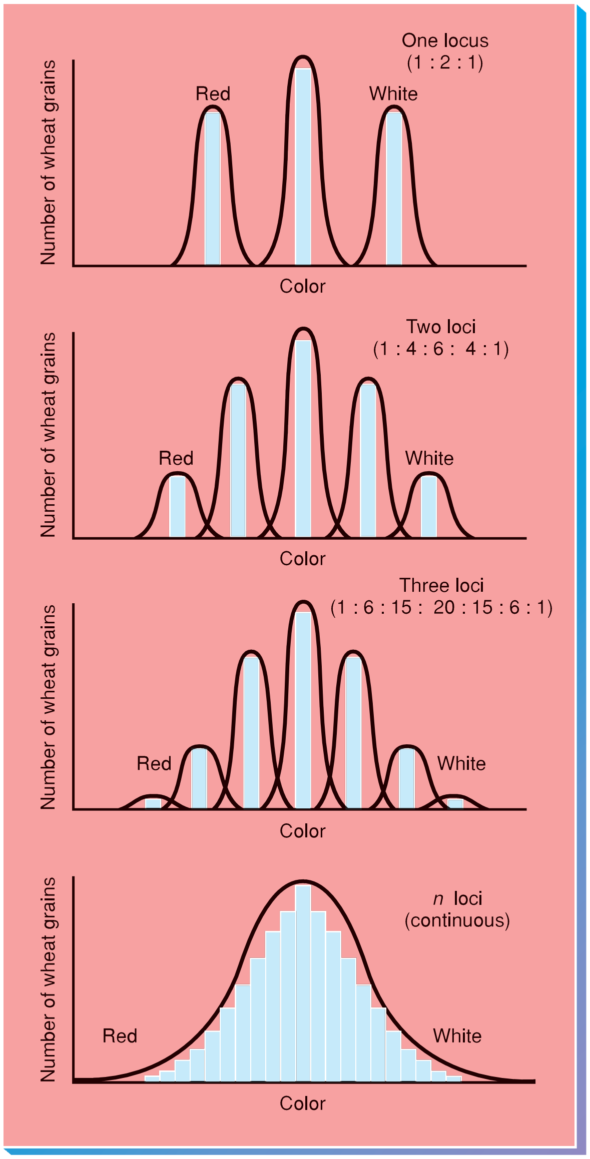


**Figure 18.5** The change in shape of the distribution as increasing numbers of independent loci control grain color in wheat. If each locus is segregating two alleles, with each allele affecting the same trait, eventually a continuous distribution will be generated in the $F_2$ generation.

different color classes down to the level of the eye's ability to perceive differences in wavelengths of light. In fact, we rapidly lose the ability to assign unique color classes to genotypes because the variation within each genotype soon causes the phenotypes to overlap. For example, with three loci, a color somewhat lighter than medium dark red may belong to the medium-dark-red class with three color alleles, or it may belong to the medium-red class with only two color alleles (fig. 18.5).

The variation within each genotype is due to the environment—that is, two organisms with the same genotype may not necessarily be identical in color because nutrition, physiological state, and many other variables influence the phenotype. Figure 18.6 shows that it is possible for the environment to obscure genotypes even in a one-locus, two-allele system. That is, a height of 17 cm could result in the $F_2$ from either the *aa* or *Aa* genotype in the figure when there is excessive variation (fig. 18.6, column 3). In the other two cases in figure 18.6, there would be virtually no organisms 17 cm tall. Systems such as those we are considering, in which each allele contributes a small unit to the phenotype, are easily influenced by the environment, with the result that the distribution of phenotypes approaches the bell-shaped curve seen at the bottom of figure 18.5.

Thus, phenotypes determined by multiple loci with alleles that contribute dosages to the phenotype will approach a continuous distribution. This type of trait is said to exhibit **continuous, quantitative,** or **metrical variation.** The inheritance pattern is **polygenic** or **quantitative.** The system is termed an **additive model** because each allele adds a certain amount to the phenotype.

From the three wheat examples just discussed, we can generalize to systems with more than three polygenic loci, each segregating two alleles. From table 18.1, we can predict the distribution of genotypes and phenotypes expected from an additive model with any number of unlinked loci segregating two alleles each. This table is useful when we seek to estimate how many loci are producing a quantitative trait, assuming it is possible to distinguish the various phenotypic classes. For example, when a strain of heavy mice was crossed with a lighter strain, the $F_1$ were of intermediate weight. When these $F_1$ were interbred, a continuous distribution of adult weights appeared in the $F_2$ generation. Since only about one mouse in 250 was as heavy as the heavy parent stock, we could guess that if an additive model holds, then four loci are segregating. This is because we expect $1/(4)^n$ to be as extreme as either parent; one in 250 is roughly $1/(4)^4 = 1/256$.

## Location of Polygenes

The fact that traits with continuous variation can be controlled by genes dispersed over the whole genome was shown by James Crow, who studied DDT resistance in *Drosophila.* A DDT-resistant strain of flies was created by growing them on increasing concentrations of the insecticide. Crow then systematically tested each chromosome for the amount of resistance it conferred. Susceptible flies were mated with resistant flies, and the sons from this cross were backcrossed. Offspring were

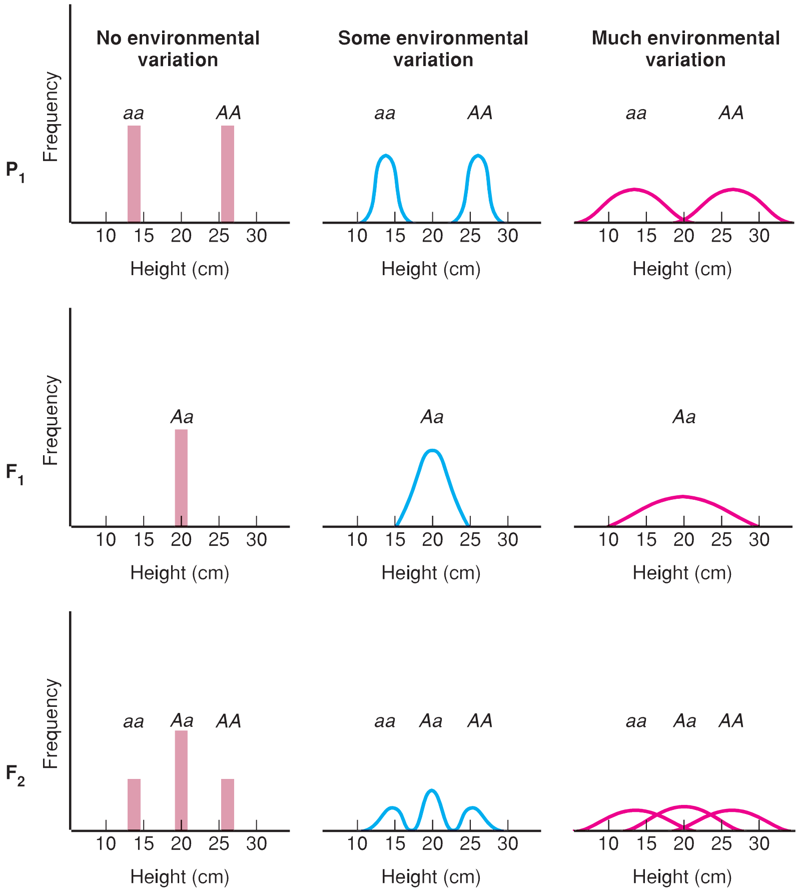
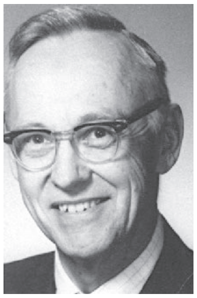


**Figure 18.6** Influence of environment on phenotypic distributions.



James F. Crow (1916– ).
(Courtesy of Dr. James F. Crow.)

then scored for the particular resistant chromosomes they contained (each chromosome had a visible marker) and were tested for their resistance to DDT. Sons were used in the backcross because there is no crossing over in males. Therefore, the sons would pass resistant and susceptible chromosomes on intact. Crow's results are shown in figure 18.7. As you can see, each chromosome has the potential to increase the fly's resistance to DDT. Thus, each chromosome contains loci (polygenes) that contribute to the phenotype of this additive trait (box 18.1).

## Significance of Polygenic Inheritance

The concept of additive traits is of great importance to genetic theory because it demonstrates that Mendelian rules of inheritance can explain traits that have a continuous distribution—that is, Mendel's rules for discrete characteristics also hold for quantitative traits. Additive traits are also of practical interest. Many agricultural products, both plant and animal, exhibit polygenic inheritance, including milk production and fruit and vegetable yield. In addition, many human traits, such as height and IQ, appear to be polygenic, although with substantial environmental components.

Historically, the study of quantitative traits began before the rediscovery of Mendel's work at the turn of the century. In fact, biologists in the early part of this century debated as to whether the "Mendelians" were correct or whether the "biometricians" were correct in regard to

**Table 18.1** **Generalities from an Additive Model of Polygenic Inheritance**

| | One Locus | Two Loci | Three Loci | *n* Loci |
|---|---|---|---|---|
| Number of gamete types produced by an $F_1$ multihybrid | 2<br>(*A, a*) | 4<br>(*AB, Ab, aB, ab*) | 8<br>(*ABC, ABc, AbC, Abc, aBC, aBc, abC, abc*) | $2^n$ |
| Number of different $F_2$ genotypes | 3<br>(*AA, Aa, aa*) | 9<br>(*AABB, AABb, AAbb, AaBB, AaBb, Aabb, aaBB, aaBb, aabb*) | 27<br>(*AABBCC, AABBCc, AABBcc, AABbCC, AABbCc, AABbcc, AAbbCC, AAbbCc, AAbbcc, AaBBCC, AaBBCc, AaBBcc, AaBbCC, AaBbCc, AaBbcc, AabbCC, AabbCc, Aabbcc, aaBBCC, aaBBCc, aaBBcc, aaBbCC, aaBbCc, aaBbcc, aabbCC, aabbCc, aabbcc*) | $3^n$ |
| Number of different $F_2$ phenotypes | 3 | 5 | 7 | $2n + 1$ |
| Number of $F_2$ as extreme as one parent or the other | 1/4<br>(*AA* or *aa*) | 1/16<br>(*AABB* or *aabb*) | 1/64<br>(*AABBCC* or *aabbcc*) | $1/4^n$ |
| Distribution pattern of $F_2$ phenotypes | 1:2:1 | 1:4:6:4:1 | 1:6:15:20:15:6:1 | $(A + a)^{2n}$ |

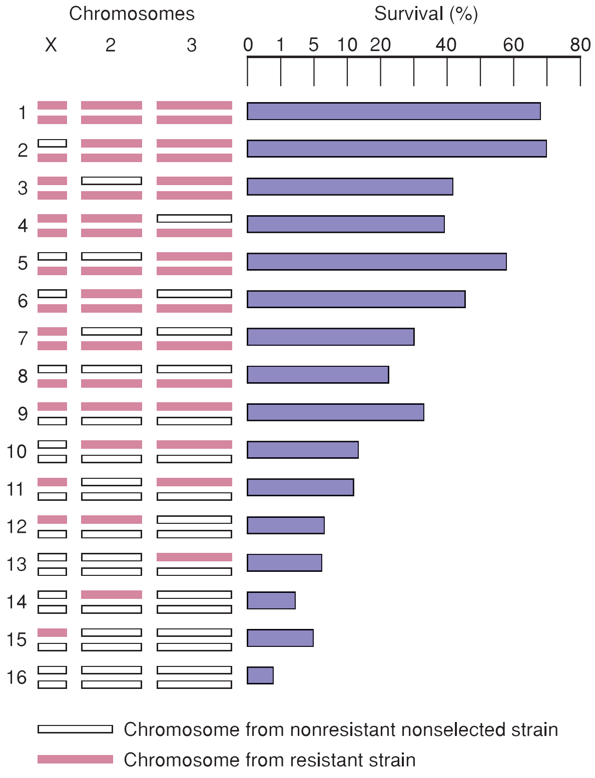


**Figure 18.7** Survival of *Drosophila* in the presence of DDT. Numbers and arrangements of DDT-resistant and susceptible chromosomes vary. (Reproduced with permission from the *Annual Review of Entomology,* Volume 2, © 1957 by Annual Reviews, Inc.)

the rules of inheritance. Biometricians used statistical techniques to study traits characterized by continuous variation and claimed that single discrete genes were not responsible for the observed inheritance patterns. They were interested in evolutionarily important facets of the phenotype—traits that can change slowly over time. Mendelians claimed that the phenotype was controlled by discrete "genes." Eventually the Mendelians were proven correct, but the biometricians' tools were the only ones suitable for studying quantitative traits.

The biometric school was founded by F. Galton and K. Pearson, who showed that many quantitative traits, such as height, were inherited. They invented the statistical tools of correlation and regression analysis in order to study the inheritance of traits that fall into smooth distributions.

## POPULATION STATISTICS

A distribution (see fig. 18.5, *bottom*) can be described in several ways. One is the formula for the shape of the curve formed by the frequencies within the distribution. A more functional description of a distribution starts by defining its center, or **mean** (fig. 18.8). As we can see from the figure, the mean is not itself enough to describe the distribution. Variation about this mean determines the actual shape of the curve. (We confine our discussion to symmetrical, bell-shaped curves called **normal distributions.** Many distributions approach a normal distribution.)

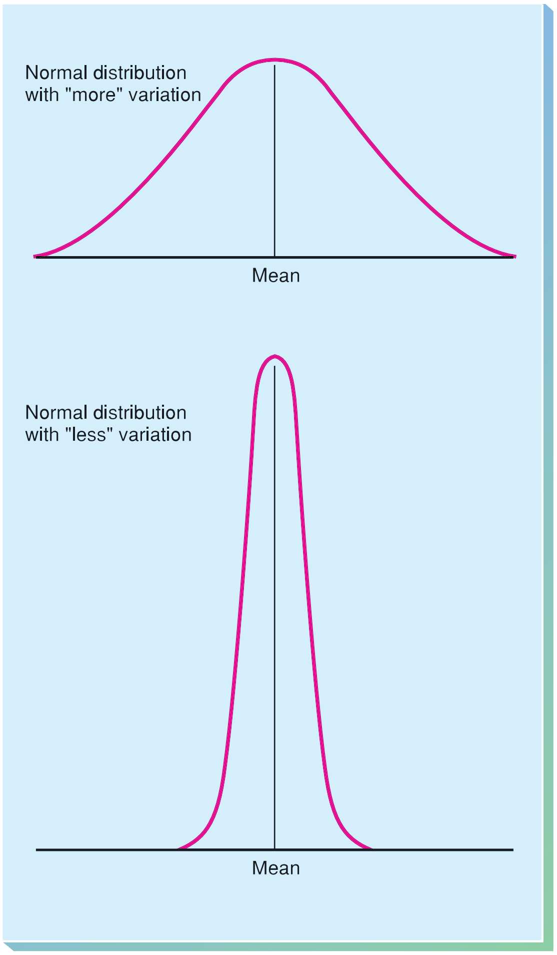


**Figure 18.8** Two normal distributions *(bell-shaped curves)* with the same mean.

## *Mean, Variance, and Standard Deviation*

The mean of a set of numbers is the arithmetic average of the numbers and is defined as

$$\bar{x} = \Sigma x/n \qquad \textbf{(18.1)}$$

in which

$\bar{x}$ = the mean

$\Sigma x$ = the summation of all values

$n$ = the number of values summed

In table 18.2, the mean is calculated for the distribution shown in figure 18.9. The variation about the mean is calculated as the average squared deviation from the mean:

$$s^2 = V = \frac{\Sigma(x - \bar{x})^2}{n - 1} \qquad \textbf{(18.2)}$$

**Table 18.2** **Hypothetical Data Set of Ear Lengths ($x$) Obtained When Corn Is Grown from an Ear of Length 11 cm**

| $x$ | $(x - \bar{x})$ | $(x - \bar{x})^2$ |
|---|---|---|
| 7 | −4.12 | 16.97 |
| 8 | −3.12 | 9.73 |
| 9 | −2.12 | 4.49 |
| 9 | −2.12 | 4.49 |
| 10 | −1.12 | 1.25 |
| 10 | −1.12 | 1.25 |
| 10 | −1.12 | 1.25 |
| 10 | −1.12 | 1.25 |
| 10 | −1.12 | 1.25 |
| 10 | −1.12 | 1.25 |
| 11 | −0.12 | 0.01 |
| 11 | −0.12 | 0.01 |
| 11 | −0.12 | 0.01 |
| 11 | −0.12 | 0.01 |
| 11 | −0.12 | 0.01 |
| 11 | −0.12 | 0.01 |
| 12 | 0.88 | 0.77 |
| 12 | 0.88 | 0.77 |
| 12 | 0.88 | 0.77 |
| 13 | 1.88 | 3.53 |
| 13 | 1.88 | 3.53 |
| 13 | 1.88 | 3.53 |
| 14 | 2.88 | 8.29 |
| 14 | 2.88 | 8.29 |
| 16 | 4.88 | 23.81 |
| $\Sigma x = 278$ | | $\Sigma(x - \bar{x})^2 = 96.53$ |

$$n = 25$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{278}{25} = 11.12$$

$$s^2 = V = \frac{\Sigma(x - \bar{x})^2}{n - 1} = \frac{96.53}{24} = 4.02$$

$$s = \sqrt{s^2} = \sqrt{4.02} = 2.0$$

This value ($V$ or $s^2$) is called the **variance.** Observe that the flatter the distribution is, the greater the variance will be.

The variance is one of the simplest measures we can calculate of variation about the mean. You might wonder why we simply don't calculate an average deviation from the mean rather than an average squared deviation. For example, we could calculate a measure of variation as

$$\frac{\Sigma(x - \bar{x})}{n - 1}$$

## BOX 18.1

# Experimental Methods

## *Mapping Quantitative Trait Loci*

Mapping the location of a standard locus is conceptually relatively easy, as we saw in the mapping of the fruit fly genome. We look for associations of phenotypes that don't segregate with simple Mendelian ratios and then map the distance between loci by the proportion of recombinant offspring. However, with quantitative loci we have a problem: We can't do simple mapping because genes contributing to the phenotype are often located across the genome. Thus, a particular continuous phenotype will be controlled by loci linked to numerous other loci, many unlinked to each other. However, with the advent of molecular techniques, it has become feasible to map polygenes.

In chapter 13, we showed how a locus can be discovered and mapped in the human genome (and other genomes) by association with molecular markers. That is, as the Human Genome Project has progressed, we have discovered restriction fragment length polymorphisms (RFLPs) that mark every region of all the chromosomes. Conceptually, there is not much difference between finding the gene for cystic fibrosis and finding the gene that contributes to a quantitative trait.

In theory, we look at a population of organisms and note various RFLPs or other molecular markers. We then look for the association of a marker and a quantitative trait. If an association exists, we can gain confidence that one or more of the polygenes controlling the trait is located in the chromosomal region near the marker. The closer the polygenes are to the markers, the more reliable our estimates are, because they depend on few crossovers taking place in that population. With many crossovers, the association between a particular marker and a particular effect diminishes. Since we don't know immediately from this method whether the region of interest has one or more polygenic loci, a new term has been coined to indicate that ambiguity. Instead of talking about polygenic loci directly, we talk of **quantitative trait loci.**

For example, consider the search for polygenes associated with geotactic behavior in fruit flies (see fig. 18.13). As selection proceeds, flies in the high and low lines diverge in their geotactic scores. The lines are also becoming homozygous for many loci since only a few parents are chosen to begin each new generation (see chapter 19). Thus, quantitative trait loci can become associated with different molecular markers in each line (fig. 1). If flies from each line are crossed, heterozygotes will be produced of both the markers and the quantitative trait loci. If there is very little crossing over between the two, three classes of $F_2$ offspring will be produced. These offspring can be grouped according to their RFLPs and then tested for their geotactic scores. If, as figure 1 suggests, a relationship exists between a locus influencing geotactic score and an RFLP,

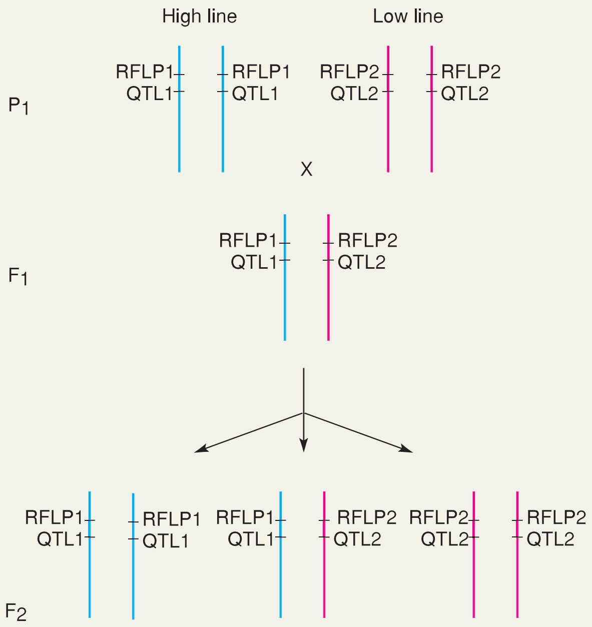


**Figure 1** Mapping a quantitative trait locus (QTL) to a particular chromosomal region using a restriction fragment length polymorphism (RFLP) marker. A hypothetical chromosome pair in the fruit fly is shown. The flies have been selected for a geotactic score; QTL1 is the locus in the high line, and QTL2 is the locus in the low line. RFLP1 is homozygous in the high line and RFLP2 is homozygous in the low line.

*continued*

**BOX 18.1 CONTINUED**

then the three groups will have different geotactic scores. We can then conclude that the region of the chromosome that contains the RFLP also contains a quantitative trait locus. Finding the right RFLP is, of course, a tedious and time-consuming task.

In a recent summary of the literature, Steven Tanksley reported that numerous quantitative trait loci have been mapped in tomatoes, corn, and other organisms. For example, five quantitative trait loci have been mapped in tomatoes for fruit growth, and eleven quantitative trait loci have been mapped in corn for plant height. Enough data seem to be present to recognize an interesting generality. That is, our definition of an additive model may need to be rethought because it appears that in almost every case studied so far, one or more of the quantitative trait loci account for a major portion of the phenotype, whereas most of the loci had very small effects. Thus, the additive model that assumes that all polygenes contribute equally to the phenotype may be wrong. However, additive models that allow different loci to contribute different degrees to the phenotype are still supported.

Also of value from locating quantitative trait loci is a new ability to estimate the number of loci affecting a quantitative trait. In this chapter, we use an estimate of extreme $F_2$ offspring to estimate the number of polygenes. There are other methods, including sophisticated statistical methods, that we will not develop here. Mapping quantitative trait loci gives us a third method, that is, simply counting the number of quantitative trait loci mapped.

As the methods of mapping quantitative trait loci have been developed, they have also been refined. High-resolution techniques under development will help us determine whether quantitative trait loci are, in fact, individual polygenes or clusters of polygenes.

(We will get to why we use $n - 1$ rather than $n$ in the denominator in a moment.) Note, however, that the above measure is zero. By the definition of the mean, the absolute value of the sum of deviations above it is equal to the absolute value of the sum of deviations below it—one is negative and the other is positive. However, by squaring each deviation, as in equation 18.2, we create a relatively simple index—the variance—which is not zero and has useful properties related to the normal distribution.

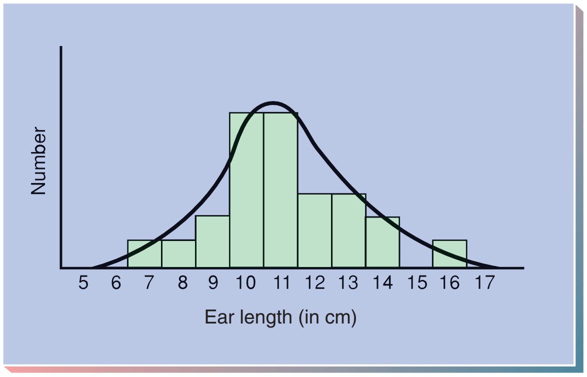


**Figure 18.9** Normal distribution of ear lengths in corn. Data are given in table 18.2.

The ear lengths measured in table 18.2 are a sample of all ear lengths in the theoretically infinite population of ears in that variety of corn. Statisticians call sample values **statistics** (and use letters from the Roman alphabet to represent them), whereas they call population values **parameters** (and use Greek letters for them). The sample value is an estimate of the true value for the population. Thus, in the variance formula (equation 18.2), the sample value, $V$ or $s^2$, is an estimate of the population variance, $\sigma^2$. When sample values are used to estimate parameters, one degree of freedom is lost for each parameter estimated. To determine the sample variance, we divide not by the sample size, but by the degrees of freedom ($n - 1$ in this case, as defined in chapter 4). The variance for the entire population (assuming we know the population mean, $\mu$, and all the data values) would be calculated by dividing by $n$. The sample variance is calculated in table 18.2.

The variance has several interesting properties, not the least of which is the fact that it is additive. That is, if we can determine how much a given variable contributes to the total variance, we can subtract that amount of variance from the total, and the remainder is caused by whatever other variables (and their interactions) affect the trait. This property makes the variance extremely important in quantitative genetic theory.

The **standard deviation** is also a measure of variation of a distribution. It is the square root of the variance:

$$s = \sqrt{V} \quad \textbf{(18.3)}$$

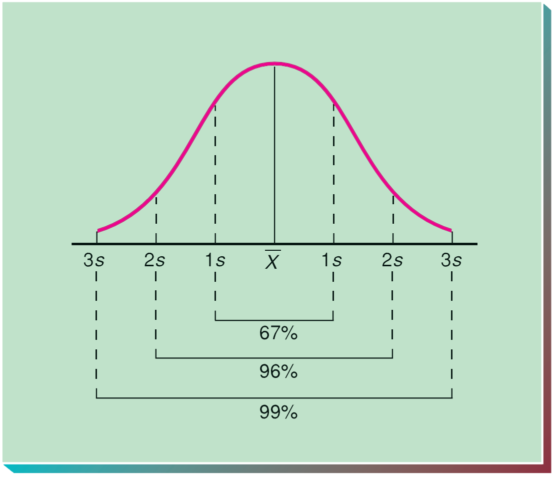


**Figure 18.10** Area under the bell-shaped curve. The abscissa is in units of standard deviation (s) around the mean ($\bar{x}$).

In a normal distribution, approximately 67% of the area of the curve lies within one standard deviation on either side of the mean, 96% lies within two standard deviations, and 99% lies within three standard deviations (fig. 18.10). Thus, for the data in table 18.2, about two-thirds of the population would have ear lengths between 9.12 and 13.12 cm (mean ± standard deviation).

One final measure of variation about the mean is the **standard error of the mean** (SE):

$$\mathrm{SE} = s/\sqrt{n}$$

The standard error (of the mean) is the standard deviation about the mean of a distribution of sample means. In other words, if we repeated the experiment many times, each time we would generate a mean value. We could then use these mean values as our data points. We would expect the variation among a population of means to be less than among individual values, and it is. Data are often summarized as "the mean ± SE." In our example of table 18.2, SE = $2.0/\sqrt{25} = 2.0/5.0 = 0.4$. We can summarize the data set of table 18.2 as 11.1 ± 0.4 (mean ± SE).

## Covariance, Correlation, and Regression

It is often desirable in genetic studies to know whether a relationship exists between two given characteristics in a series of individuals. For example, is there a relationship between height of a plant and its weight, or between scholastic aptitude and grades, or between a phenotypic measure in parents and their offspring? If one increases, does the other also? An example appears in table 18.3; the same data set is graphed in figure 18.11, in what is referred to as a scatter plot. A relation does appear between the two variables. With increasing wing length in midparent (the average of the two parents: x-axis), there is an increase in offspring wing length (y-axis). We can determine how closely the two variables are related by calculating a **correlation coefficient**—an index that goes from −1.0 to +1.0, depending on the degree of relationship between the variables. If there is no relation (if the variables are independent), then the correlation coefficient will be zero. If there is perfect correlation, where an increase in one variable is associated with a proportional increase in the other, the coefficient will be +1.0. If an increase in one is associated with a proportional decrease in the other, the coefficient will be −1.0 (fig. 18.12). The formula for the correlation coefficient ($r$) is

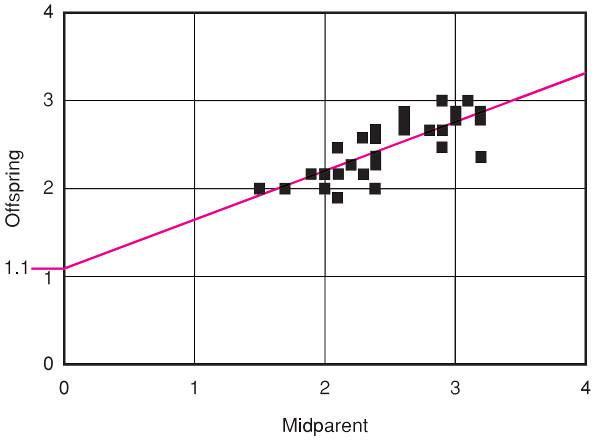


**Figure 18.11** The relationship between two variables, parental and offspring wing length in fruit flies, measured in millimeters. *Midparent* refers to the average wing length of the two parents. The line is the statistical regression line. (Source: Data from D. S. Falconer, *Introduction to Quantitative Genetics,* 2d ed. [London: Longman, 1981].)

$$r = \frac{\text{covariance of } x \text{ and } y}{s_x \cdot s_y} \tag{18.4}$$

where $s_x$ and $s_y$ are the standard deviations of $x$ and $y$, respectively.

To calculate the correlation coefficient, we need to define and calculate the **covariance** of the two variables, cov($x$, $y$). The covariance is analogous to the variance, but it involves the simultaneous deviations from the means of both the $x$ and $y$ variables:

$$\mathrm{cov}(x, y) = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{n - 1} \tag{18.5}$$

The analogy between variance and covariance can be seen by comparing equations 18.5 and 18.2. The variances, standard deviations, and covariance are calculated in table 18.3, in which the correlation coefficient, $r$, is 0.78. (There are computational formulas available that substantially cut down on the difficulty of calculating these statistics. If a computer or calculator is used, only the individual data points need to be entered—most computers and many calculators can be programmed to do all the computations.)

Many experiments deal with a situation in which we assume that one variable is dependent on the other (in a cause-and-effect relationship). For example, we may ask, what is the relationship of DDT resistance in *Drosophila* to an increased number of DDT-resistant alleles? With more of these alleles (see fig. 18.7), the DDT resistance of the flies should increase. Number of DDT-resistant alleles is the independent variable, and resistance of the flies is the dependent variable. That is, a fly's resistance is dependent on the number of DDT-resistant alleles it has, not the other way around. Going back to figure 18.11, we could make the assumption that offspring wing length is dependent on parental wing length. If this were so, a technique called *regression analysis* could be used. This analysis allows us to predict an offspring's wing length ($y$ variable) given a particular midparental wing length ($x$ variable). (It is important to note that regression analysis assumes a cause-and-effect relationship, whereas correlation analysis does not.)

The formula for the straight-line relationship (regression line) between the two variables is $y = a + bx$, where $b$ is the slope of the line (change in $y$ divided by change in $x$, or $\Delta y/\Delta x$) and $a$ is the $y$-intercept of the line (see fig. 18.11). To define any line, we need only to calculate the slope, $b$, and the $y$ intercept, $a$:

$$b = \text{cov}(x, y)/s_x^2 \qquad \textbf{(18.6)}$$

$$a = \bar{y} - b\bar{x} \qquad \textbf{(18.7)}$$

Thus equipped, if a cause-and-effect relationship does exist between the two variables, we can predict a $y$ value given any $x$ value. We can either use the formula $y = a + bx$ or graph the regression line and directly determine the $y$ value for any $x$ value. We now continue our examination of the genetics of quantitative traits.

**Table 18.3** **The Relationship Between Two Variables, $x$ and $y$ ($x$ = the midparent—average of the two parents—in wing length in fruit flies in millimeters; $y$ = the offspring measurement)**

| $x$ | $y$ | $x$ | $y$ | $x$ | $y$ | $x$ | $y$ |
|---|---|---|---|---|---|---|---|
| 1.5 | 2 | 2.2 | 2.3 | 2.4 | 2.7 | 2.9 | 2.7 |
| 1.7 | 2 | 2.3 | 2.2 | 2.4 | 2.7 | 2.9 | 2.7 |
| 1.9 | 2.2 | 2.3 | 2.6 | 2.6 | 2.7 | 2.9 | 3 |
| 2 | 2 | 2.4 | 2 | 2.6 | 2.7 | 3 | 2.8 |
| 2 | 2.2 | 2.4 | 2.3 | 2.6 | 2.8 | 3 | 2.8 |
| 2 | 2.2 | 2.4 | 2.4 | 2.6 | 2.9 | 3 | 2.9 |
| 2.1 | 1.9 | 2.4 | 2.6 | 2.8 | 2.7 | 3.1 | 3 |
| 2.1 | 2.2 | 2.4 | 2.6 | 2.8 | 2.7 | 3.2 | 2.4 |
| 2.1 | 2.5 | 2.4 | 2.6 | 2.9 | 2.5 | 3.2 | 2.8 |
| | | | | | | 3.2 | 2.9 |

$\Sigma x = 92.7 \qquad n = 37 \qquad \Sigma y = 93.2$

$$\bar{x} = \frac{\Sigma x}{n} = 2.51 \qquad \bar{y} = \frac{\Sigma y}{n} = 2.52$$

$$s_x^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} = 0.19 \qquad s_y^2 = \frac{\Sigma(y - \bar{y})^2}{n - 1} = 0.10$$

$$s_x = \sqrt{s_x^2} = 0.44 \qquad s_y = \sqrt{s_y^2} = 0.32$$

$$\text{cov}(x, y) = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{n - 1} = 0.11$$

$$r = \frac{\text{cov}(x, y)}{s_x s_y} = \frac{0.11}{(0.44)(0.32)} = 0.78$$

*Source:* Data from D. S. Falconer, *Introduction to Quantitative Genetics*, 2d ed. (London: Longman, 1981).

*Note:* Data are graphed in figure 18.11.

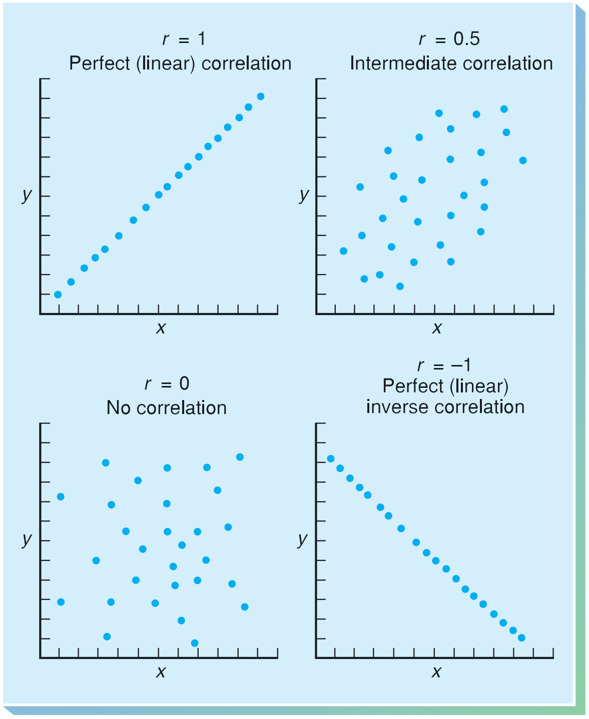


**Figure 18.12** Plots showing varying degrees of correlation within data sets.

**Table 18.4** **Johannsen's Findings of Relationship Between Bean Weights of Parents and Their Progeny**

| Weight of Parent Beans | Weight of Progeny Beans (centigrams) | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | n | Mean ± SE |
| 65–75 | | | | 2 | 3 | 16 | 37 | 71 | 104 | 105 | 75 | 45 | 19 | 12 | 3 | 2 | 494 | 58.47 ± 0.43 |
| 55–65 | | | 1 | 9 | 14 | 51 | 79 | 103 | 127 | 102 | 66 | 34 | 12 | 6 | 5 | | 609 | 54.37 ± 0.41 |
| 45–55 | | | 4 | 20 | 37 | 101 | 204 | 287 | 234 | 120 | 76 | 34 | 17 | 3 | 1 | | 1,138 | 51.45 ± 0.27 |
| 35–45 | 5 | 6 | 11 | 36 | 139 | 278 | 498 | 584 | 372 | 213 | 69 | 20 | 4 | 3 | | | 2,238 | 48.62 ± 0.18 |
| 25–35 | | 2 | 13 | 37 | 58 | 133 | 189 | 195 | 115 | 71 | 20 | 2 | | | | | 835 | 46.83 ± 0.30 |
| 15–25 | | | 1 | 3 | 12 | 29 | 61 | 38 | 25 | 11 | | | | | | | 180 | 46.53 ± 0.52 |
| Totals | 5 | 8 | 30 | 107 | 263 | 608 | 1,068 | 1,278 | 977 | 622 | 306 | 135 | 52 | 24 | 9 | 2 | 5,491 | 50.39 ± 0.13 |

## POLYGENIC INHERITANCE IN BEANS

In 1909, W. Johannsen, who studied seed weight in the dwarf bean plant (*Phaseolus vulgaris*), demonstrated that polygenic traits are controlled by many genes. The parent population was made up of seeds (beans) with a continuous distribution of weights. Johannsen divided this parental group into classes according to weight, planted them, self-fertilized the plants that grew, and weighed the $F_1$ beans. He found that the parents with the heaviest beans produced the progeny with the heaviest beans, and the parents with the lightest beans produced the progeny with the lightest beans (table 18.4). There was a significant correlation coefficient between parent and progeny bean weight ($r = 0.34 \pm 0.01$). He continued this work by beginning nineteen lines (populations) with beans from various points on the original distribution and selfing each successive generation for the next several years. After a few generations, the means and variances stabilized within each line. That is, when Johannsen chose, within each line, parent plants with heavier-than-average or lighter-than-average seeds, the offspring had the parental mean with the parental variance for seed size. For example, in one line, plants with both the lightest average bean weights (24 centigrams) and plants with the heaviest average bean weights (47 cg) produced offspring with average bean weights of 37 cg. By selfing the plants each generation, Johannsen had made them more and more homozygous, thus lowering the number of segregating polygenes. Therefore, the lines became homozygous for certain of the polygenes (different in each line), and any variation in bean weight was then caused only by the environment. Johannsen thus showed that quantitative traits were under the control of many segregating loci.

## SELECTION EXPERIMENTS

Selection experiments are done for several reasons. Plant and animal breeders select the most desirable individuals as parents in order to improve their stock. Population geneticists select specific characteristics for study in order to understand the nature of quantitative genetic control.

For example, *Drosophila* were tested in a fifteen-choice maze for geotactic response (fig. 18.13). The maze was on its side, so at every intersection, a fly had to make a choice between going up or going down. The flies with the highest scores were chosen as parents for the "high" line (positive geotaxis; favored downward direction), and the flies with the lowest score were chosen as parents for the "low" line (negative geotaxis; favored upward direction). The same selection was made for each generation. As time progressed, the two lines diverged quite significantly. This tells us that there is a large genetic component to the response; the experimenters are successfully amassing more of the "downward" alleles in the high line and more of the "upward" alleles in the low line. Several other points emerge from this graph. First, the high and low responses are slightly different, or asymmetrical. The high line responded more quickly, leveled out more quickly, and tended toward the original state more slowly after selection was relaxed. (The relaxation of selection occurred when the parents were a random sample of the adults rather than the extremes for geotactic scores.) The low line responded more slowly and erratically. In addition, the low line returned toward the original state more quickly when selection was relaxed.

The nature of these responses (fig. 18.13) indicates that the high line became more homozygous than the low line. This is shown by the former's response when selection is relaxed: It has exhausted a good deal of its variability for the polygenes responsible for geotaxis. The low line,

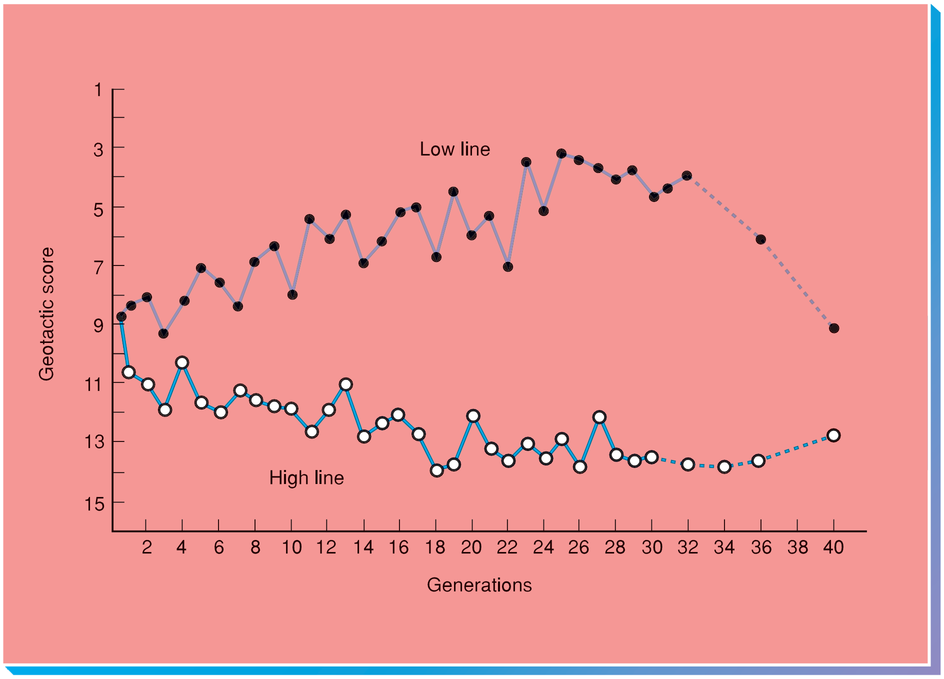


**Figure 18.13** Selection for geotaxis. The *dotted lines* represent relaxed selection. (Source: Data from T. Dobzhansky and B. Spassky, "Artificial and natural selection for two behavioral traits in *Drosophila pseudoobscura,*" *Proceedings of the National Academy of Sciences, USA,* 62:75–80, 1969.)

however, seems to have much of its original genetic variability, because the relaxation of selection caused the mean score of this line to increase rapidly. It still had enough genetic variability to head back to the original population mean. The response to a selection experiment is one way that plant and animal breeders can predict future response.

## HERITABILITY

Plant and animal breeders want to improve the yields of their crops to the greatest degree they can. They must choose the parents of the next generation on the basis of this generation's yields; thus, they are continually performing selection experiments. Breeders run into two economic problems. They cannot pick only the very best to be the next generation's parents because (1) they cannot afford to decrease the size of a crop by using only a very few select parents and (2) they must avoid **inbreeding depression,** which occurs when plants are self-fertilized or animals are bred with close relatives for many generations. After frequent inbreeding, too much homozygosity occurs, and many genes that are slightly or partially deleterious begin to show themselves, depressing vigor and yield. (Chapter 19 presents more on inbreeding.) Thus, breeders need some index of the potential response to selection so that they can then get the greatest amount of selection with the lowest risk of inbreeding depression.

### Realized Heritability

Breeders often calculate a **heritability** estimate, a value that predicts to what extent their selection will be successful. Heritability is defined in the following equation:

$$H = \frac{Y_O - \overline{Y}}{Y_P - \overline{Y}} = \frac{\text{gain}}{\text{selection differential}} \qquad \textbf{(18.8)}$$

in which

$H$ = heritability
$Y_O$ = offspring yield
$\overline{Y}$ = mean yield of the population
$Y_P$ = parental yield

From this equation, we can see that heritability is the gain in yield divided by the amount of selection practiced (fig. 18.14). $Y_O - \overline{Y}$ is the improvement over the population average due to $Y_P - \overline{Y}$, which is the amount of difference between the parents and the population average. If there is no gain ($Y_O = \overline{Y}$), then the heritability

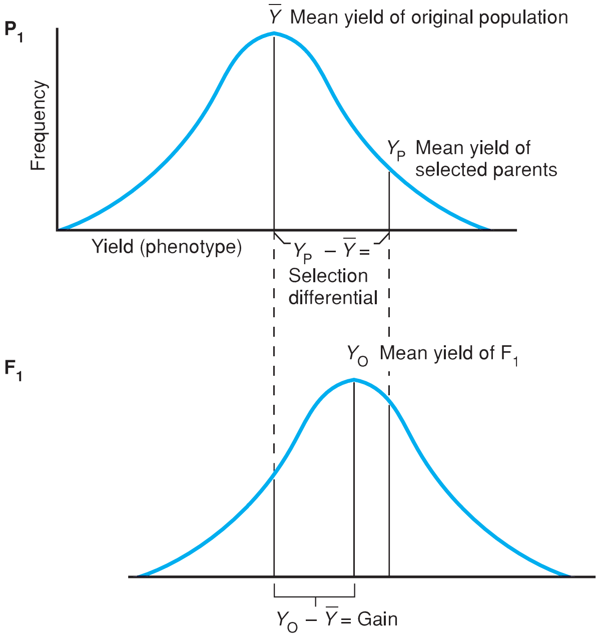


**Figure 18.14** Realized heritability is the gain in yield divided by the selection differential when offspring are produced by parents with a mean yield that differs from that of the general population.

**Table 18.5** **Some Realized Heritabilities**

| Animal | Trait | Heritability |
|---|---|---|
| Cattle | Birth weight | 0.49 |
| | Milk yield | 0.30 |
| Poultry | Body weight | 0.31 |
| | Egg production | 0.30 |
| | Egg weight | 0.60 |
| Swine | Birth weight | 0.06 |
| | Growth rate | 0.30 |
| | Litter size | 0.15 |
| Sheep | Wool length | 0.55 |
| | Fleece weight | 0.40 |

will be zero, and breeders will know that no matter how much selection they practice, they will not improve their crops and might as well not waste their time. Since this value is calculated after the breeding has been done, it is referred to as **realized heritability.** Some typical values for realized heritabilities are shown in table 18.5.

The following example may help to clarify the calculation of realized heritability. The number of bristles on the sternopleurite, a thoracic plate in *Drosophila,* is under polygenic control. In a population of flies, the mean bristle number was 6.4. Three pairs of flies served as parents; they had a mean of 7.2 bristles. Their offspring had a mean of 6.6 bristles. Hence, $Y_O = 6.6$, $\bar{Y} = 6.4$, and $Y_P = 7.2$. Dividing the gain by the selection differential—that is, substituting in equation 18.8—gives us

$$H = \frac{6.6 - 6.4}{7.2 - 6.4} = \frac{0.2}{0.8} = 0.25$$

If both a low line and a high line were begun, and if both were carried over several generations, the heritability would be measured by the final difference in means of the high and low lines (gain) divided by the cumulative selection differentials summed for both the high and low lines.

Note from figure 18.13 that the response to selection declines with time as the selected population becomes homozygous for various alleles controlling the trait. As the response declines, the calculated heritability value itself declines. After intense or prolonged selection, heritability may be zero. It does not mean that the trait is not controlled by genes, only that there is no longer a response to selection. Hence, heritability is specific for a particular population at a particular time. Intense selection exhausts the genetic variability, rendering the response to selection, and thus the heritability itself, zero.

Quantitative geneticists treat the realized heritability as an estimate of **true heritability.** True heritability is actually viewed in two different ways: as *heritability in the narrow sense* and *heritability in the broad sense.* We define these on the basis of partitioning of the variance of the quantitative character under study.

## *Partitioning of the Variance*

Given that the variance of a distribution has genetic and environmental causes, and given that the variance is additive, we can construct the following formula:

$$V_{\mathrm{Ph}} = V_{\mathrm{G}} + V_{\mathrm{E}} \tag{18.9}$$

in which

$V_{\mathrm{Ph}}$ = total phenotypic variance
$V_{\mathrm{G}}$ = variance due to genotype
$V_{\mathrm{E}}$ = variance due to environment

Throughout the rest of this discussion, we will stay with this model. We could construct a more complex variance model if there are interactions between variables. For example, if one genotype responded better in one soil condition than in another soil condition, this environment-genotype interaction would require a separate variance term ($V_{\mathrm{GE}}$).

The variance due to the genotype ($V_{\mathrm{G}}$) can be further broken down according to the effects of additive polygenes ($V_{\mathrm{A}}$), dominance ($V_{\mathrm{D}}$), and epistasis ($V_{\mathrm{I}}$) to give us a final formula:

$$V_{\mathrm{Ph}} = V_{\mathrm{A}} + V_{\mathrm{D}} + V_{\mathrm{I}} + V_{\mathrm{E}} \tag{18.10}$$

We can now define the two commonly used—and often confused—measures of heritability. *Heritability in the narrow sense* is

$$H_N = V_A/V_{Ph} \tag{18.11}$$

This heritability is the proportion of the total phenotypic variance caused by additive genetic effects. It is the heritability of most interest to plant and animal breeders because it predicts the magnitude of the response under selection.

*Heritability in the broad sense* is

$$H_B = V_G/V_{Ph} \tag{18.12}$$

This heritability is the proportion of the total phenotypic variance caused by all genetic factors, not just additive factors. It measures the extent to which individual differences in a population are caused by genetic differences. This measure is the one most often used by psychologists. We are concerned primarily with $H_N$, heritability in the narrow sense.

## Measurement of Heritability

Three general methods are used to estimate heritability. First, as discussed earlier, we can measure heritability by the response of a population to selection. Second, we can directly estimate the components of variance by minimizing one component; the remaining variance can then be attributed to other causes. For example, by minimizing environmental causes of variance, we can estimate the genetic component directly. Or, by eliminating the genetic causes of variance, we can estimate the environmental component directly. Third, we can measure the similarity between relatives. We look now at the latter two methods.

Variance components can be minimized in several different ways. If we use genetically identical organisms, then the additive, dominance, and epistatic variances are zero, and all that is left is the environmental variance. For example, F. Robertson determined the variance components for the length of the thorax in *Drosophila*. The total variance ($V_{Ph}$) in a genetically heterogeneous population was 0.366 (measured directly from the distribution of the trait, as in tables 18.2 and 18.3). He then looked at the variance in flies that were genetically homogeneous. These were from isolated lines inbred in the laboratory over many generations to become virtually homozygous. Robertson studied the $F_1$ in several different matings of inbred lines and found the variance in thorax length to be 0.186 ($V_E$). By subtraction (0.366 − 0.186), we know that the total genetic variance ($V_G$) was 0.180. From this, we can calculate heritability in the broad sense as

$$H_B = V_G/V_{Ph} = 0.180/0.366 = 0.49$$

To calculate a heritability in the narrow sense, it is necessary to extract the components of the genetic variance, $V_G$.

Genetic variance can be measured directly by minimizing the influence of the environment. This is most easily done with plants grown in a greenhouse. Under that circumstance, environmental variables, such as soil quality, water, and sunlight, can be controlled to a very high degree. Hence, the variance among individuals grown under these circumstances is almost all genetic variance. The total phenotypic variance can be obtained from the plants grown under natural circumstances. This allows us to calculate heritability in the broad sense.

Several methods exist to sort out the additive from the dominant and epistatic portions of the genetic variance. The methods rely mostly on correlations between relatives. That is, the expected amount of genetic similarity between certain relatives can be compared with the actual similarity. The expected amount of genetic similarity is the proportion of genes shared; this is a known quantity for any form of relatedness. For example, parents and offspring have half their genes in common. The relation of observed and expected correlations between relatives is a direct measure of heritability in the narrow sense. We can thus define

$$H_N = r_{obs}/r_{exp} \tag{18.13}$$

in which $r_{obs}$ is the observed correlation between the relatives, and $r_{exp}$ is the expected correlation. The expected correlation is simply the proportion of the genes in common.

We must point out that the observed correlation between relatives can be artificially inflated if the environments are not random. Since we know that relatives frequently share similar (or correlated) environments, they may show a phenotypic similarity irrespective of genetic causes. It is important to keep that in mind, especially when we analyze human traits, where it may be almost impossible to rule out or quantify environmental similarity. Hence, $r_{obs}$ may be inflated, which will inflate $H_N$.

In human beings, finger-ridge counts (fingerprints, fig. 18.15) have a very high heritability; there seems to be very little environmental interference in the embryonic development of the ridges (table 18.6). Monozygotic twins are from the same egg, which divides into two embryos at a very early stage. They have identical genotypes. Dizygotic twins result from the simultaneous fertilization of two eggs. They have the same genetic relationship as siblings. (However, environmental influences may be different; they may be treated differently by relatives and friends.) The data therefore suggest that human finger ridges are almost completely controlled by additive genes with a negligible input from environmental and dominance variation. Few human traits are controlled this simply (table 18.7).

This brief discussion should make it clear that the components of the total variance can be estimated. For a given quantitative trait, the total variance can be mea-

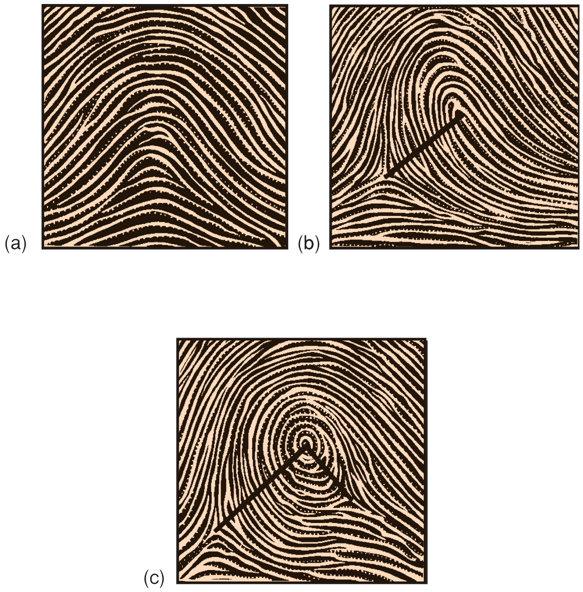


**Figure 18.15** The three basic fingerprint patterns. Ridges are counted where they intersect the line connecting a triradius with a loop or whorl center. (*a*) An arch; there is no triradius; the ridge count is zero. (*b*) A loop; thirteen ridges. (*c*) A whorl; there are two triradii and counts of seventeen and eight (the higher one is routinely used). (From Sarah B. Holt, "Quantitative genetics of finger-print patterns," *British Medical Bulletin,* 17. Copyright © 1961 Churchill Livingstone Medical Journals, Edinburgh, Scotland. Reprinted by permission.)

**Table 18.6** **Correlations Between Relatives, and Heritabilities, for Finger-Ridge Counts**

| Relationship | $r_{obs}$ | $r_{exp}$ | $H_N$ |
|---|---|---|---|
| Mother-child | 0.48 | 0.50 | 0.96 |
| Father-child | 0.49 | 0.50 | 0.98 |
| Siblings | 0.50 | 0.50 | 1.00 |
| Dizygotic twins | 0.49 | 0.50 | 0.98 |
| Monozygotic twins | 0.95 | 1.00 | 0.95 |

*Source:* From Sarah B. Holt, "Quantitative genetics of finger-print patterns," *British Medical Bulletin,* 17. Copyright © 1961 Churchill Livingstone Medical Journals, Edinburgh, Scotland. Reprinted by permission.

sured directly. If identical genotypes can be used, then the environmental component of variance can be determined. By correlation of various relatives, it is possible to directly measure heritability in the narrow sense. If heritability is known, and if the total phenotypic variance is known, then all that are left, assuming no interaction, are the dominance and epistatic components. In practice, the epistatic components are usually ignored. Thus, operationally, all that is left is the dominance variance, obtained by subtracting the additive from the total genetic variance. In addition, plant and animal breeders use sophisticated statistical techniques of covariance and variance analysis, techniques that are beyond our scope.

**Table 18.7** **Some Estimates of Heritabilities ($H_N$) for Human Traits and Disorders**

| Trait | Heritability |
|---|---|
| Schizophrenia | 0.85 |
| Diabetes Mellitus | |
| Early onset | 0.35 |
| Late onset | 0.70 |
| Asthma | 0.80 |
| Cleft Lip | 0.76 |
| Heart Disease, Congenital | 0.35 |
| Peptic Ulcer | 0.37 |
| Depression | 0.45 |
| Stature* | 1.00+ |

* A heritability higher than one can be obtained when the correlation among relatives is higher than expected. This is usually the result of dominant alleles.

## QUANTITATIVE INHERITANCE IN HUMAN BEINGS

As with most human studies, the measurement of heritability is limited by a lack of certain types of information. We cannot develop pure human lines, nor can we manipulate human beings into various kinds of environments or do selection experiments. However, certain kinds of information are available that allow some estimation of heritabilities.

### *Skin Color*

Skin color is a quantitative human trait for which a simple analysis can be done on naturally occurring matings. Certain groups of people have black skin; other groups do not. Many of these groups breed true in the sense that skin colors stay the same generation after generation within a group; when groups intermarry and produce offspring, the $F_1$ are intermediate in skin color. In turn, when $F_1$ individuals intermarry and produce offspring, the skin color of the $F_2$ is, on the average, about the same as the $F_1$, but with more variation (fig. 18.16). The data are consistent with a model of four loci, each segregating two alleles. At each locus, one allele adds a measure of color, whereas the other adds none.

## *IQ and Other Traits*

In human beings, twin studies have been helpful in estimating the heritability of quantitative traits. One way of looking at quantitative traits is by the **concordance** among twins. Concordance means that if one twin has the trait, the other does also. Discordance means one has the trait and the other does not. Table 18.8 shows some concordance values. High concordance of monozygotic as compared with dizygotic twins is another indicator of the heritability of a trait. Concordance values for measles susceptibility and handedness, which are similar for both monozygotic and dizygotic twins, demonstrate the environmental influence on some traits.

Some monozygotic twins (MZ) have been reared apart. The same is true for dizygotic twins (DZ) and nontwin siblings. IQ (intelligence quotient) is a measure of intelligence highly correlated among relatives, indicating a strong genetic component. In three studies of monozygotic twins reared apart, the average correlation in IQ was 0.72. In thirty-four studies of monozygotic twins reared together, the average correlation in IQ was 0.86; dizygotic twins reared together have an average correlation of 0.60 in IQ. Thus, it is clear that there is a genetic influence on IQ. However, experts disagree strongly on the environmental role in shaping IQ and the exact meaning of IQ as a functional measure of intelligence (box 18.2).

At present, twin studies are emerging from the shadow of a scandal involving a knighted British psychologist, Cyril Burt (1883–1971), who did classical twin research on the inheritance of IQ. Burt was posthumously accused of fraud, an accusation that was almost universally accepted and that cast doubt on all of his data and conclusions. More recently, new information seemed to cast doubt on the charges of fraud. These on-again, off-again charges have been a focus of scientific interest.

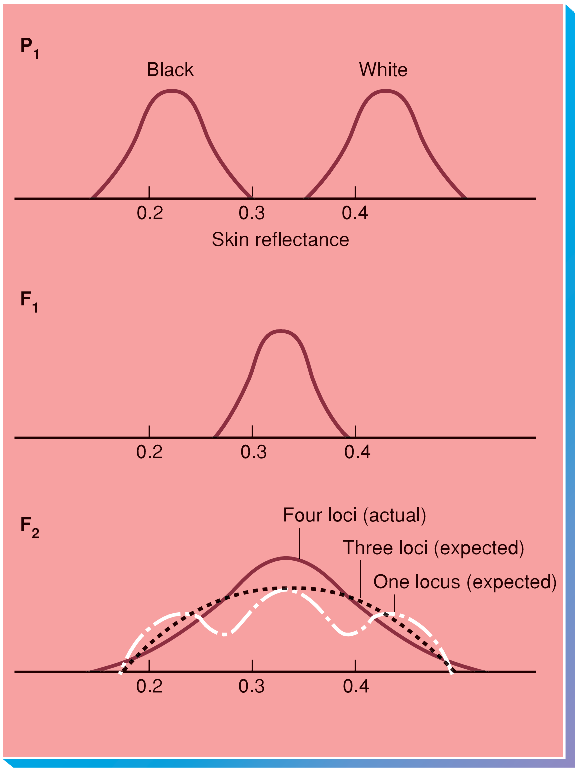


**Figure 18.16** Inheritance of skin color in human beings. Four loci are probably involved.

**Table 18.8** **Concordance of Traits Between Identical and Fraternal Twins**

| | Identical (MZ) Twins (%) | Fraternal (DZ) Twins (%) |
|---|---|---|
| Hair color | 89 | 22 |
| Eye color | 99.6 | 28 |
| Blood pressure | 63 | 36 |
| Handedness (left or right) | 79 | 77 |
| Measles | 95 | 87 |
| Clubfoot | 23 | 2 |
| Tuberculosis | 53 | 22 |
| Mammary cancer | 6 | 3 |
| Schizophrenia | 80 | 13 |
| Down syndrome | 89 | 7 |
| Spina bifida | 72 | 33 |
| Manic-depression | 80 | 20 |

## BOX 18.2

## Ethics and Genetics

### Human Behavioral Genetics

The study of human behavioral genetics was at first associated with the **eugenics** movement, founded in the late nineteenth century by Francis Galton, one of the founders of quantitative genetics. Eugenics was a movement designed to improve humanity by better breeding. This movement was tainted by bad science done by people with strong prejudices. However, although still controversial, the study of human behavioral genetics is back in vogue. The political climate has changed, and scientific methods to study human behaviors have improved. Even some of the strongest critics against these studies have changed their minds when confronted with the discovery of particular behavioral genes through the mapping of quantitative trait loci. In addition, advocates for people with many human conditions, such as homosexuality and mental illness, feel that if these traits are shown to be genetic in origin, then those who have them will be treated as people with medical conditions rather than as social outcasts.

Even with better methods and more objective practitioners, the study of human behavioral genetics is still difficult to achieve. Some traits are very poorly defined and may be complex mixtures of phenotypes, such as schizophrenia. Other traits are just difficult to define, such as alcoholism, criminal tendency, and aggressiveness. To make things more complicated, several recent studies that seemed to isolate genes for specific behavioral traits were not verified or were later retracted. In one case, a gene for manic-depressive behavior was isolated in an Amish population. However, when the study was expanded, new cases were discovered that were not linked to the particular marker locus. The result was that a "found" genetic locus was lost. To their credit, the workers were quick to retract their conclusions.

Currently, there are intriguing results suggesting that divorce, aggression, and dyslexia are under genetic influence. For example, in a recent study, investigators measured a heritability of 0.52 for divorce. This doesn't mean that "divorce genes" exist, but rather that genes for certain personality traits might predispose a person to divorce. We should make it clear that genetic control does not mean that the environment does not play a role in these traits, just that there are genes that are influential also, sometimes very significantly.

In retrospect, it should not be surprising that genes influence much of our behavior. There are numerous animal studies confirming genetic control of behaviors, indicating that the same would be found in people. As long as the research is done in a competent fashion and the results are not "politicized," human behavior genetics should not only be a reasonable area of study, but an exciting one as we learn more about ourselves.

## SUMMARY

**STUDY OBJECTIVE 1:** To understand the patterns of inheritance of phenotypic traits controlled by many loci 531–535

Some genetically controlled phenotypes do not fall into discrete categories. This type of variation is referred to as quantitative, continuous, or metrical variation. The genetic control of this variation is referred to as polygenic control. If the number of controlling loci is small, and offspring fall into recognizable classes, it is possible to analyze the genetic control of the phenotypes with standard methods. Polygenes controlling DDT resistance are located on all chromosomes in *Drosophila.*

**STUDY OBJECTIVE 2:** To investigate the way that geneticists and statisticians describe and analyze normal distributions of phenotypes 535–542

When phenotypes fall into a continuous distribution, the methods of genetic analysis change. We must describe a distribution using means, variances, and standard deviations. Then we must describe the relationship between two variables using variances and correlation coefficients.

**STUDY OBJECTIVE 3:** To define and measure heritability, the unit of inheritance of variation in traits controlled by many loci 542–547

Equipped with statistical tools, we analyzed the genetic control of continuous traits. The heritability estimates tell us how much of the variation in the distribution of a trait can be attributed to genetic causes. Heritability in the narrow sense is the relative amount of variance due to additive loci. Heritability in the broad sense is the relative amount of variance due to all genetic components, including dominance and epistasis. In practice, heritability can be calculated as realized heritability—gain divided by selection differential. Estimates of human heritabilities can be constructed from correlations among relatives, concordance and discordance between twins, and studies of monozygotic twins reared apart.

## SOLVED PROBLEMS

**PROBLEM 1:** In a certain stock of wheat, grain color is controlled by four loci acting according to an additive model. How many different gametes can a tetrahybrid produce? How many different genotypes will result if tetrahybrids are self-fertilized? What will be the phenotypic distribution of these genotypes?

***Answer:*** Assume the *A*, *B*, *C*, and *D* loci with *A* and *a*, *B* and *b*, *C* and *c*, and *D* and *d* alleles, respectively. A tetrahybrid will have the genotype *Aa Bb Cc Dd.* A gamete can get either allele at each of four independently assorting loci, so there are $2^4 = 16$ different gametes. Three genotypes are possible for each locus, two homozygotes and a heterozygote. Therefore, for four independent loci, there are $3^4 = 81$ different genotypes. Phenotypes are distributed according to the binomial distribution. Thus, there will be a pattern of $(A + a)^{2n} = (A + a)^8$; a ratio of 1:8:28:56:70:56:28:8:1 of phenotypes with decreasing red color from left to right, eight red colors plus white.

**PROBLEM 2:** In horses, white facial markings are inherited in an additive fashion. These markings are scored on a scale that begins at zero. In a particular population, the average score is 2.2. A group of horses with an average score of 3.4 is selected to be parents of the next generation. The offspring of this group of selected parents have a mean score of 3.1. What is the realized heritability of white facial markings in this herd of horses?

***Answer:*** This is a simple selection experiment; the data fit our equation for realized heritability (equation 18.8). In this case:

$Y_O$ = offspring yield = 3.1
$\overline{Y}$ = mean yield of the population = 2.2
$Y_P$ = parental yield = 3.4

Substituting into equation 18.8:

$$H = \frac{Y_O - \overline{Y}}{Y_P - \overline{Y}} = \frac{3.1 - 2.2}{3.4 - 2.2} = \frac{0.9}{1.2} = 0.75$$

**PROBLEM 3:** Corn growing in a field in Indiana had a lysine (amino acid) content of 2.0%, with a variance of 0.16. When grown in the greenhouse under controlled and uniform conditions, the mean lysine content was again 2.0%, but the variance was 0.09. What measure of heritability can you calculate?

***Answer:*** We use equation 18.9 for the calculation of heritability by partitioning of the variance ($V_{\text{Ph}} = V_{\text{G}} + V_{\text{E}}$). In this case:

$V_{\text{Ph}}$ = total phenotypic variance = 0.16
$V_{\text{G}}$ = variance due to genotype = 0.09
$V_{\text{E}}$ = variance due to environment = ?

In the greenhouse, we have minimized environmental variance, meaning the total genotypic variance = 0.09. If we subtract this from the total variance, we get the original environmental variance: $0.16 - 0.09 = 0.7$. Heritability in the broad sense is the genetic variance divided by the total phenotypic variance, or $0.09/0.16 = 0.56$.

# EXERCISES AND PROBLEMS[*]

**TRAITS CONTROLLED BY MANY LOCI**

**1.** A variety of squash has fruits that weigh about 5 pounds each. In a second variety, the average weight is 2 pounds. When the two varieties are crossed, the $F_1$ produce fruit with an average weight of 3.5 pounds. When two of these are crossed, their offspring produce a range of fruit weights, from 2 to 5 pounds. Of two hundred offspring, three produce fruits weighing about 5 pounds and three produce fruits about 2 pounds in weight. Approximately how many allelic pairs are involved in the weight difference between the varieties, and approximately how much does each effective gene contribute to the weight?

**2.** In rabbit variety 1, ear length averages 4 inches. In a second variety, it is 2 inches. Hybrids between the varieties average 3 inches in ear length. When these hybrids are crossed among themselves, the offspring exhibit a much greater variation in ear length, ranging from 2 to 4 inches. Of five hundred $F_2$ animals, two have ears about 4 inches long, and two have ears about 2 inches long. Approximately how many allelic pairs are involved in determining ear length, and how much does each effective gene seem to contribute to the length of the ear? What do the distributions of $P_1$, $F_1$, and $F_2$ probably look like?

**3.** Assume that height in people depends on four pairs of alleles. How can two persons of moderate height produce children who are much taller than they are? Assume that the environment is exerting a negligible effect.

**4.** How do polygenes differ from traditional Mendelian genes?

**5.** If skin color is caused by additive genes, can matings between individuals with intermediate-colored skin produce light-skinned offspring? Can such matings produce dark-skinned offspring? Can matings between individuals with light skin produce dark-skinned offspring? (*See also* QUANTITATIVE INHERITANCE IN HUMAN BEINGS)

**6.** The tabulated data from Emerson and East ("The Inheritance of Quantitative Characters in Maize," 1913, *Univ. Nebraska Agric. Exp. Sta. Bull,* no. 2) show the results of crosses between two varieties of corn and their $F_2$ offspring (see the table on ear length in corn). Provide an explanation for these data in terms of number of allelic pairs controlling ear length. Do all the genes involved affect length additively? Explain.

**7.** In *Drosophila,* a marker strain exists containing dominant alleles that are lethal in the homozygous condition on both chromosome 2, 3, and 4 homologues. These six lethal alleles are within inversions, so there is virtually no crossing over. The strain thus remains perpetually heterozygous for all six loci and therefore all three chromosome pairs. (Geneticists use a shorthand notation in these "balanced-lethal" systems in which only the dominant alleles on a chromosome are shown, with a slash separating the two homologous chromosomes.) The markers are: chromosome 2, Curly and Plum (*Cy/Pm,* shorthand for $CyPm^{+}/Cy^{+}Pm$); chromosome 3, Hairless and stubble (*H/S*); and chromosome 4, Cell and Minute(4) (*Ce/M*[4]). With this strain, which allows you to follow particular chromosomes by the presence or absence of phenotypic markers, construct crosses to give the strains Crow used (see fig. 18.7) to determine the location of polygenes for DDT resistance.

**8.** A red-flowered plant is crossed with a yellow-flowered plant to produce $F_1$ plants with orange flowers. The $F_1$ offspring are selfed, and they produce plants with flowers in a range of seven different colors. How many genes are probably involved in color production?

**Ear Length in Corn (cm)**

| | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variety $P_{60}$ | 4 | 21 | 24 | 8 | | | | | | | | | | | | | |
| Variety $P_{54}$ | | | | | | | | | 3 | 11 | 12 | 15 | 26 | 15 | 10 | 7 | 2 |
| $F_1$ | | | | | 1 | 12 | 12 | 14 | 17 | 9 | 4 | | | | | | |
| $F_2(F_1 \times F_1)$ | | | 1 | 10 | 19 | 26 | 47 | 73 | 68 | 68 | 39 | 25 | 15 | 9 | 1 | | |

---

[*] Answers to selected exercises and problems are on page A-20.

**9.** A plant with a genotype of *aabb* and a height of 40 cm is crossed with a plant with a genotype of *AABB* and a height of 60 cm. If each dominant allele contributes to height additively, what is the expected height of the $F_1$ progeny?

**10.** If the $F_1$ generation in the cross in problem 9 is selfed, what proportion of the $F_2$ offspring would you expect to be 50 cm tall?

**11.** Two strains of wheat were compared for the time required to mature. Strain X required fourteen days, and strain Y required twenty-eight days. The strains were crossed, and the $F_1$ generation was selfed. One hundred $F_2$ progeny out of 6,200,000 matured in fourteen days or less. How many genes may be involved in maturation?

### POPULATION STATISTICS

**12.** A geneticist wished to know if variation in the number of egg follicles produced by chickens was inherited. As a first step in his experiments, he wished to determine if the number of eggs laid could be used to predict the number of follicles. If this were true, he could then avoid killing the chickens to obtain the data he needed. He obtained the following data from fourteen chickens.

| Chicken Number | Eggs Laid | Ovulated Follicles |
|---|---|---|
| 1 | 39 | 37 |
| 2 | 29 | 34 |
| 3 | 46 | 52 |
| 4 | 28 | 26 |
| 5 | 31 | 32 |
| 6 | 25 | 25 |
| 7 | 49 | 55 |
| 8 | 57 | 65 |
| 9 | 51 | 44 |
| 10 | 21 | 25 |
| 11 | 42 | 45 |
| 12 | 38 | 26 |
| 13 | 34 | 29 |
| 14 | 47 | 30 |

Calculate a correlation coefficient. Graph the data, and then calculate the slope and $y$-intercept of the regression line. Draw the regression line on the same graph.

**13.** The following table (data from Ehrman and Parsons, 1976, *The Genetics of Behavior,* 121, Sunderland, Mass.: Sinauer Associates) gives heights in centimeters of eleven pairs of brothers and sisters. Calculate a correlation coefficient and a heritability. Is this realized heritability, heritability in the broad sense, or heritability in the narrow sense?

| Pair | Brother | Sister | Pair | Brother | Sister |
|---|---|---|---|---|---|
| 1 | 180 | 175 | 7 | 178 | 165 |
| 2 | 173 | 162 | 8 | 186 | 163 |
| 3 | 168 | 165 | 9 | 183 | 168 |
| 4 | 170 | 160 | 10 | 165 | 160 |
| 5 | 178 | 165 | 11 | 168 | 157 |
| 6 | 180 | 157 | | | |

How can environmental factors influence this heritability value? (*See also* HERITABILITY)

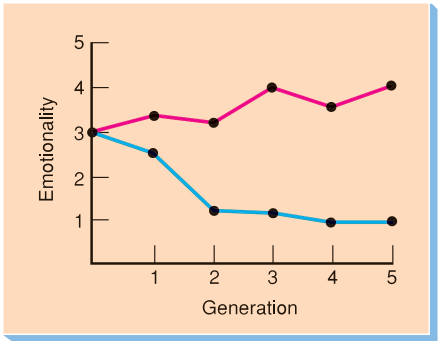
**14.** You determine the following variance components for leaf width in a particular species of plant:

| | |
|---|---|
| Additive genetic variance ($V_A$) | 4.0 |
| Dominance genetic variance ($V_D$) | 1.8 |
| Epistatic variance ($V_I$) | 0.5 |
| Environmental variance ($V_E$) | 2.5 |

Calculate the broad sense and narrow sense heritabilities. (*See also* HERITABILITY)

### SELECTION EXPERIMENTS

**15.** Psychologists refer to defecation rate in rats as "emotionality." The data shown in the accompanying figure (data modified from Broadhurst, 1960, *Experiments in Personality,* vol. 1, London: Eysenck) show mean emotionality scores during five generations in high and low selection lines. In the final generation, the parental mean was 4 for the high line and 0.9 for the low line. The cumulative selection differential is 5 for each line. Calculate realized heritability overall, and separate heritabilities for each line. Do these differ? Why? Why was the response to selection asymmetrical? (*See also* HERITABILITY)



**16.** Data were gathered during a selection experiment for six-week body weight in mice. Graph these data and calculate a realized heritability. (*See also* HERITABILITY)

| | **High Line** | | | **Low Line** | | |
|---|---|---|---|---|---|---|
| **Generation** | $\bar{Y}$ | $Y_P$ | $Y_O$ | $\bar{Y}$ | $Y_P$ | $Y_O$ |
| 0 | 21 | | | 21 | | |
| 1 | | 24 | 22 | | 18 | 20 |
| 2 | | 24 | 23 | | 18 | 20 |
| 3 | | 26 | 23 | | 18 | 20 |
| 4 | | 26 | 24 | | 16 | 19 |
| 5 | | 26 | 23 | | 16 | 18 |

## HERITABILITY

**17.** Outstanding athletic ability is often found in several members of a family. Devise a study to determine to what extent athletic ability is inherited. (What is "outstanding athletic ability"?)

**18.** Variations in stature are almost entirely due to heredity. Yet average height has increased substantially since the Middle Ages, and the increase in the height of children of immigrants to the United States, as compared with the height of the immigrants themselves, is especially noteworthy. How can these observations be reconciled?

**19.** Would you expect good nutrition to increase or decrease the heritability of height?

**20.** Two adult plants of a particular species have extreme phenotypes for height (1 foot tall and 5 feet tall), a quantitative trait. If you had only one uniformly lighted greenhouse, how would you determine whether the variation in plant height is environmentally or genetically determined? How would you attempt to estimate the number of allelic pairs that may be involved in controlling this trait?

**21.** The components of variance for two characters of *D. melanogaster* are shown in the following table (data from A. Robertson, "Optimum Group Size in Progeny Testing," *Biometrics,* 13:442–50, 1957). Estimate the dominance and epistatic components, and calculate heritabilities in the narrow and broad sense.

| **Variance Components** | **Thorax Length** | **Eggs Laid in Four Days** |
|---|---|---|
| $V_{Ph}$ | 100 | 100 |
| $V_A$ | 43 | 18 |
| $V_E$ | 51 | 38 |
| $V_D + V_I$ | ? | ? |

**22.** In a mouse population, the average tail length is 10 cm. Six mice with an average tail length of 15 cm are interbred. The mean tail length in their progeny is 13.5 cm. What is the realized heritability?

**23.** The narrow sense heritability of egg weight in chickens in one coop is 0.5. A farmer selects for heavier eggs by breeding a few chickens with heavier eggs. He finds a difference of 9 g in the mean egg weights of selected and unselected chickens. By how much can he expect egg weight to increase in the selected chickens?

**24.** If, in a population of swine, the narrow sense heritability of maturation weight is 0.15, the phenotypic variance is 100 $lb^2$, the total genetic variance is 50 $lb^2$, and the epistatic variance is 0, calculate the dominance genetic variance and the environmental variance.

**25.** A group of four-month-old hogs has an average weight of 170 pounds. The average weight of selected breeders is 185 pounds. If the heritability of weight is 40%, what is the expected average weight of the first generation progeny?

## QUANTITATIVE INHERITANCE IN HUMAN BEINGS

**26.** Does schizophrenia seem to have a strong genetic component (see table 18.8)? Explain.

## CRITICAL THINKING QUESTIONS

**1.** Several cases mentioned in the text reported and then retracted the discovery of human genes controlling specific traits. Barring fraud, what might cause a scientist to retract a study of this type?

**2.** Monozygotic twins share identical genes. Under what conditions could they show discordance of traits?

*Suggested Readings for chapter 18 are on page B-19.*

# 19

# POPULATION GENETICS

## *The Hardy-Weinberg Equilibrium and Mating Systems*

### STUDY OBJECTIVES

1. To understand the concept of population-level genetic processes 553
2. To learn the assumptions and nature of the Hardy-Weinberg equilibrium and its extensions 554
3. To test whether a population is in Hardy-Weinberg equilibrium 557
4. To analyze the process and consequences of nonrandom mating in diploid populations 560

### STUDY OUTLINE

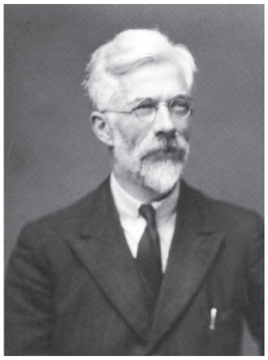
**Hardy-Weinberg Equilibrium** 553
    Calculating Allelic Frequencies 553
    Assumptions of Hardy-Weinberg Equilibrium 554
    Proof of Hardy-Weinberg Equilibrium 555
    Generation Time 556
    Testing for Fit to Hardy-Weinberg Equilibrium 557
**Extensions of Hardy-Weinberg Equilibrium** 558
    Multiple Alleles 558
    Multiple Loci 559
**Nonrandom Mating** 560
    Inbreeding 560
    Pedigree Analysis 562
    Population Analysis 564
*Summary* 566
*Solved Problems* 566
*Exercises and Problems* 567
*Critical Thinking Questions* 569
**Box 19.1 The Determination of Lethal Equivalents** 562

The cheetah (*Acinonyx jubatus*) is in peril of extinction; it has very low genetic variability.

(Gregory G. Dimijian, MD/Photo Researchers, Inc.)

Evolution is a process that takes place in populations of organisms. To study evolution, we need to shift our focus to population genetics, the algebraic description of the genetic makeup of a population and the changes in allelic frequencies in populations over time. This chapter is the first of three that looks at what population genetics can tell us about the way evolution proceeds.

Almost all of the mathematical foundations of genetic changes in populations were developed in a short period of time during the 1920s and 1930s by three men: R. A. Fisher, J. B. S. Haldane, and S. Wright. Some measure of disagreement emerged among these men, but they disagreed on which evolutionary processes were more important, not on how the processes worked. Since the 1960s, excitement has arisen in the field of population genetics, primarily on three fronts. First, the high-speed computer has made it possible to do a large amount of arithmetic in a very short period of time; thus, complex simulations of real populations can be added to the repertoire of the experimental geneticist. Second, electrophoresis has provided a means of gathering the large amount of empirical data necessary to check some of the assumptions used in mathematical models. The information and interpretation of the electrophoretic data have generated some controversy about the role of "neutral" evolutionary changes in natural populations. Last, newer techniques of molecular genetics are being used to analyze the relationships among species and the rate of evolutionary processes. We consider these studies later.



Sir Ronald A. Fisher (1890–1962). (Courtesy of The National Portrait Gallery, England.)

Sewall Wright (1889–1988). (Courtesy of Dr. Sewall Wright.)

## HARDY-WEINBERG EQUILIBRIUM

Let us begin with a few definitions. For the most part, we define a *species* as a group of organisms potentially capable of interbreeding. Most species are made up of **populations,** interbreeding groups of organisms that are usually subdivided into partially isolated breeding groups called **demes.** As we will see, it is these demes, or local populations, that can evolve.

In 1908, G. H. Hardy, a British mathematician, and W. Weinberg, a German physician, independently discovered a rule that relates allelic and genotypic frequencies in a population of diploid, sexually reproducing individuals if that population has random mating, large size, no mutation or migration, and no selection. The rule has three aspects:

1. The allelic frequencies at an autosomal locus in a population will not change from one generation to the next (allelic-frequency equilibrium).
2. The genotypic frequencies of the population are determined in a predictable way by the allelic frequencies (genotypic-frequency equilibrium).
3. The equilibrium is neutral. That is, if it is perturbed, it will be reestablished within one generation of random mating at the new allelic frequencies (if all the other requirements are maintained).

### Calculating Allelic Frequencies

If we consider an autosomal locus in a diploid, sexually reproducing species, allelic frequencies can be measured in either of two ways. The first is simply by counting genes:

$$\text{frequency of the } a \text{ allele, } q, = \frac{\text{number of } a \text{ alleles}}{\text{total number of alleles}}$$

The expression "frequency of" can be shortened to *f*( ). For example, the frequency of the *a* allele is written as *f*(*a*). Since the homozygotes have two of a given allele and heterozygotes have only one, and since the total number of alleles is twice the number of individuals (each individual carries two alleles), we can calculate allelic frequencies in the following manner. Consider, for example, the phenotypic distribution of MN blood types (controlled by the codominant *M* and *N* alleles) among two hundred persons chosen randomly in Columbus, Ohio:

$$\begin{aligned} &\text{type M } (MM \text{ genotype}) &= 114 \\ &\text{type MN } (MN \text{ genotype}) &= 76 \\ &\text{type N } (NN \text{ genotype}) &= \underline{10} \\ & & 200 \end{aligned}$$

Then,

$$p = f(M) = \frac{2(114) + 76}{2(200)} = \frac{304}{400} = 0.76$$

Similarly,

$$q = f(N) = \frac{2(10) + 76}{2(200)} = \frac{96}{400} = 0.24$$

Alternatively, because the frequencies of the two alleles, $M$ and $N$, must add up to unity ($p + q = 1$, $q = 1 - p$, and $p = 1 - q$), if we know that $p = 0.76$, then $q = 1 - 0.76 = 0.24$.

Another way of calculating allelic frequencies is based on knowledge of the genotypic frequencies. In this example, the frequencies are

$$f(MM) = \frac{114}{200} = 0.57$$

$$f(MN) = \frac{76}{200} = 0.38$$

$$f(NN) = \frac{10}{200} = 0.05$$

We derive an expression for calculating $p$ and $q$ based on genotypic frequencies as follows:

$$\begin{aligned} p = f(M) &= \frac{2 \times \text{number of } MM + \text{number of } MN}{2 \times \text{total number}} \\ &= \frac{2 \times \text{number of } MM}{2 \times \text{total number}} + \frac{\text{number of } MN}{2 \times \text{total number}} \\ &= f(MM) + (1/2)f(MN) \end{aligned}$$

and,

$$\begin{aligned} q = f(N) &= \frac{2 \times \text{number of } NN + \text{number of } MN}{2 \times \text{total number}} \\ &= \frac{2 \times \text{number of } NN}{2 \times \text{total number}} + \frac{\text{number of } MN}{2 \times \text{total number}} \\ &= f(NN) + (1/2)f(MN) \end{aligned}$$

Thus, allelic frequencies can be calculated as the frequency of homozygotes, plus half the frequency of heterozygotes, as follows:

$$\begin{aligned} p = f(M) &= f(MM) + (1/2)f(MN) \\ &= 0.57 + (1/2)0.38 = 0.76 \\ q = f(N) &= f(NN) + (1/2)\,f(MN) \\ &= 0.05 + (1/2)0.38 = 0.24 \end{aligned}$$

or

$$q = 1 - p = 1 - 0.76 = 0.24$$

Note that these two methods (counting alleles and using genotypic frequencies) are algebraically identical and thus give identical results.

## Assumptions of Hardy-Weinberg Equilibrium

We will consider a population of diploid, sexually reproducing organisms with a single autosomal locus segregating two alleles (i.e., every individual is one of three genotypes—*MM, MN*, or *NN*). Later on, we generalize the discussion to include multiple alleles and multiple loci. For the moment, the focus is on a genetic system such as the *MN* locus in human beings. The following major assumptions are necessary for the Hardy-Weinberg equilibrium to hold.

### Random Mating

The first assumption is **random mating,** which means that the probability that two genotypes will mate is the product of the frequencies (or probabilities) of the genotypes in the population. If the *MM* genotype makes up 90% of a population, then any individual has a 90% chance (probability = 0.9) of mating with a person with an *MM* genotype. The probability of an *MM* by *MM* mating is (0.9)(0.9), or 0.81.

Deviations from random mating come about for two reasons: choice or circumstance. If members of a population choose individuals of a particular phenotype as mates more or less often than at random, the population is engaged in **assortative mating.** If individuals with similar phenotypes are mating more often than at random, *positive assortative mating* is in force; if matings occur between individuals with dissimilar phenotypes more often than at random, *negative assortative mating,* or **disassortative mating,** is at work.

Deviations from random mating also arise when mating individuals are either more closely related genetically or more distantly related than individuals chosen at random from the population. **Inbreeding** is the mating of related individuals, and **outbreeding** is the mating of genetically unrelated individuals. Inbreeding is a consequence of pedigree relatedness (e.g., cousins) and small population size.

One of the first counterintuitive observations of population genetics is that deviations from random mating alter genotypic frequencies but not allelic frequencies. Envision a population in which every individual is the parent of two children. On the average, each individual will pass on one copy of each of his or her alleles. Assortative mating and inbreeding will change the zygotic (genotypic) combinations from one generation to the next, but will not change which alleles are passed into the next generation. Thus genotypic, but not allelic, frequencies change under nonrandom mating.

### *Large Population Size*

Even when an extremely large number of gametes is produced in each generation, each successive generation is the result of a sampling of a relatively small portion of the gametes of the previous generation. A sample may not be an accurate representation of a population, especially if the sample is small. Thus, the second assumption of the Hardy-Weinberg equilibrium is that the population is infinitely large. A large population produces a large sample of successful gametes. The larger the sample, the greater the probability that the allelic frequencies of the offspring will accurately represent the allelic frequencies in the parental population. When populations are small or when alleles are rare, changes in allelic frequencies take place due to chance alone. These changes are referred to as **random genetic drift,** or just *genetic drift.*

### *No Mutation or Migration*

Allelic and genotypic frequencies may change through the loss or addition of alleles through mutation or migration (immigration or emigration) of individuals from or into a population. The third and fourth assumptions of the Hardy-Weinberg equilibrium are that neither mutation nor migration causes such allelic loss or addition in the population.

### *No Natural Selection*

The final assumption necessary to the Hardy-Weinberg equilibrium is that no individual will have a reproductive advantage over another individual because of its genotype. In other words, no natural selection is occurring. (Artificial selection, as practiced by animal and plant breeders, will also perturb the Hardy-Weinberg equilibrium of captive populations.)

In summary, the Hardy-Weinberg equilibrium holds (is exactly true) for an infinitely large, randomly mating population in which mutation, migration, and natural selection do not occur. In view of these assumptions, it seems that such an equilibrium would never be characteristic of natural populations. However, this is not the case. Hardy-Weinberg equilibrium is approximated in natural populations for two major reasons. First, the consequences of violating some of the assumptions, such as no mutation or infinitely large population size, are small. Mutation rates, for example, are on the order of one change per locus per generation per $10^6$ gametes. Thus, there is virtually no measurable effect of mutation in a single generation. In addition, populations do not have to be infinitely large to act as if they were. As we will see, a relatively small population can still closely approximate Hardy-Weinberg equilibrium. In other words, minor deviations from the other assumptions can still result in a good fit to the equilibrium; only major deviations can be detected statistically. Second, the Hardy-Weinberg equilibrium is extremely resilient to change because, regardless of the perturbation, the equilibrium is usually reestablished after only one generation of random mating. The new equilibrium will be, however, at the new allelic frequencies—the Hardy-Weinberg equilibrium does not "return" to previous allelic values.

## Proof of Hardy-Weinberg Equilibrium

The three properties of the Hardy-Weinberg equilibrium are that (1) allelic frequencies do not change from generation to generation, (2) allelic frequencies determine genotypic frequencies, and (3) the equilibrium is achieved in one generation of random mating. We will concentrate for a moment on the second property. In a population of individuals segregating the *A* and *a* alleles at the *A* locus, each individual will be one of three genotypes: *AA*, *Aa*, or *aa*. If $p = f(A)$ and $q = f(a)$, then we can predict the genotypic frequencies in the next generation. If all the assumptions of the Hardy-Weinberg equilibrium are met, the three genotypes should occur in the population in the same frequencies at which gametes would be randomly drawn in pairs from a **gene pool.** A gene pool is defined as all of the alleles available among the reproductive members of a population from which gametes can be drawn. Thus,

$$f(AA) = (p \times p) = p^2$$

$$f(Aa) = (p \times q) + (q \times p) = 2pq$$

$$f(aa) = (q \times q) = q^2$$

demonstrates the second property of the Hardy-Weinberg equilibrium (fig. 19.1).

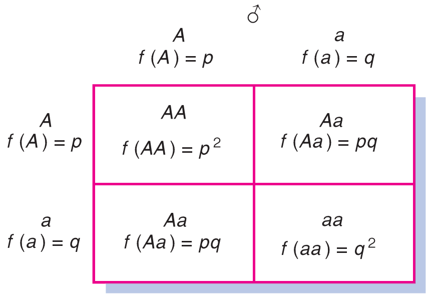


**Figure 19.1** Gene pool concept of zygote formation. Males and females have the same frequencies of the two alleles: $f(A) = p$ and $f(a) = q$. After one generation of random mating, the three genotypes, *AA, Aa,* and *aa,* have the frequencies of $p^2$, $2pq$, and $q^2$, respectively.

Another way of demonstrating the properties of the Hardy-Weinberg equilibrium for the one-locus, two-allele case in sexually reproducing diploids is by simply observing the offspring of a randomly mating, infinitely large population. Let the initial frequencies of the three genotypes be any values that sum to one; for example, let $X$, $Y$, and $Z$ be the proportions of the *AA*, *Aa*, and *aa* genotypes, respectively. The proportions of offspring after one generation of random mating are as shown in table 19.1. For example, the probability that an *AA* individual will mate with an *AA* individual is $X \times X$, or $X^2$. Since all the offspring of this mating are *AA*, they are counted only under the *AA* column of offspring in table 19.1. When all possible matings are counted, the offspring with each genotype are summed. The proportion of *AA* offspring is $X^2 + XY + (1/4)Y^2$, which factors to $(X + [1/2]Y)^2$. Recall that the frequency of an allele is the frequency of its homozygote plus half the frequency of the heterozygote. Hence, $X + (1/2)Y$ is the frequency of *A*, since $X = f(AA)$ and $Y = f(Aa)$. If $p = f(A)$, then $(X + [1/2]Y)^2$ is $p^2$. Thus, after one generation of random mating, the proportion of *AA* homozygotes is $p^2$. Similarly, the frequency of *aa* homozygotes after one generation of random mating is $Z^2 + YZ + (1/4)Y^2$, which factors to $(Z + [1/2]Y)^2$, or $q^2$. The frequency of heterozygotes when summed and factored (table 19.1) is $2(X + [1/2]Y)(Z + [1/2]Y)$, or $2pq$. Therefore, after one generation of random mating, the three genotypes (*AA*, *Aa*, and *aa*) occur as $p^2$, $2pq$, and $q^2$.

Looking at the first property of the Hardy-Weinberg equilibrium, that allelic frequencies do not change generation after generation, we can ask, Have the allelic frequencies changed from one generation to the next (from the parents to the offspring)? Before random mating, the frequency of the *A* allele is, by definition, $p$:

$$f(A) = p = f(AA) + (1/2)f(Aa) = X + (1/2)Y$$

After random mating, the frequency of the *A* homozygote is $p^2$, and the frequency of the heterozygote is $2pq$. Thus, the frequency of the *A* allele, the frequency of its homozygote plus half the frequency of the heterozygotes, is

$$\begin{aligned} f(A) &= f(AA) + (1/2)f(Aa) \\ &= p^2 + (1/2)(2pq) \\ &= p^2 + pq = p(p + q) \\ &= p \text{ (remember, } p + q = 1) \end{aligned}$$

Thus, in a randomly mating population of sexually reproducing diploid individuals, the allelic frequency, $p$, does not change from generation to generation. Here, by observing the offspring of a randomly mating population, we have proven all three properties of the Hardy-Weinberg equilibrium.

## Generation Time

Although generation interval is commonly thought of as the average age of the parents when their offspring are born, the statistical concept of a generation is more complex. Demographers use formulas that relate generation time to the age of reproducing females, the reproductive level of each age group, and the probability of survival in each age group. Here, to avoid these complexities, we will use **discrete generations,** unless otherwise noted. That is, we will assume that all the individuals drawn in a sam-

**Table 19.1 Proportions of Offspring in a Randomly Mating Population Segregating the *A* and *a* Alleles at the *A* locus: $X = f(AA)$, $Y = f(Aa)$, and $Z = f(aa)$**

| | | Offspring | | |
|---|---|---|---|---|
| **Mating** | **Proportion** | ***AA*** | ***Aa*** | ***aa*** |
| $AA \times AA$ | $X^2$ | $X^2$ | | |
| $AA \times Aa$ | $XY$ | $(1/2)XY$ | $(1/2)XY$ | |
| $AA \times aa$ | $XZ$ | | $XZ$ | |
| $Aa \times AA$ | $XY$ | $(1/2)XY$ | $(1/2)XY$ | |
| $Aa \times Aa$ | $Y^2$ | $(1/4)Y^2$ | $(1/2)Y^2$ | $(1/4)Y^2$ |
| $Aa \times aa$ | $YZ$ | | $(1/2)YZ$ | $(1/2)YZ$ |
| $aa \times AA$ | $XZ$ | | $XZ$ | |
| $aa \times Aa$ | $YZ$ | | $(1/2)YZ$ | $(1/2)YZ$ |
| $aa \times aa$ | $Z^2$ | | | $Z^2$ |
| Sum | $(X + Y + Z)^2$ | $(X + [1/2]Y)^2$ | $2(X + [1/2]Y)(Z + [1/2]Y)$ | $(Z + [1/2]Y)^2$ |

ple, for purposes of determining allelic and genotypic frequencies, are drawn from the same generation, and that, in resampling the population, the second sample represents the offspring of the first generation. The discrete-generation model holds for organisms such as annual plants and fruit flies maintained under laboratory conditions, with no breeding among individuals of different generations. Generations that overlap, as in populations of human beings and many other organisms, usually are better described by somewhat more complex mathematical models.

## Testing for Fit to Hardy-Weinberg Equilibrium

There are several ways to determine whether a given population conforms to the Hardy-Weinberg equilibrium at a particular locus. However, the question usually arises when there is just a single sample from a population, representing only one generation. Can the existence of the Hardy-Weinberg equilibrium be determined with just one sample? The answer is that we can determine whether the three genotypes (*AA*, *Aa*, and *aa*) occur with the frequencies $p^2$, $2pq$, and $q^2$. If they do, then the population is considered to be in Hardy-Weinberg proportions; if not, then the population is not considered to be in Hardy-Weinberg proportions.

### MN Blood Types

To determine whether observed and expected allelic frequencies are the same, we can use the chi-square statistical test. In a chi-square test, we compare an observed number with an expected number. In this case, the observed values are the actual numbers of the three genotypes in the sample, and the expected values come from the prediction that the genotypes will occur in the $p^2$, $2pq$, and $q^2$ proportions. An analysis for the Ohio MN blood-type data is presented in table 19.2. The agreement between observed and expected numbers is very good, obvious even before the calculation of the chi-square value. Since the critical chi-square for one degree of freedom at the 0.05 level is 3.841 (see table 4.4), we find that the Ohio population does not deviate from Hardy-Weinberg proportions at the *MN* locus.

Earlier (chapter 4), we used the chi-square statistic to test how well real data fit an expected data set based on a ratio predicted before the test. For example, we tested the data against a 3:1 ratio in table 4.2. In that case, the number of degrees of freedom was simply the number of independent categories: the total number of categories minus one. Here, however, our expected ratio is derived from the data set itself. The values $p^2$, $2pq$, and $q^2$ came from $p$ and $q$, which were estimated from the data. In this case, we lose one additional degree of freedom for every independent value we estimate from the data. If we calculate $p$ from a sample, we lose one degree of freedom. However, we do not lose a degree of freedom for estimating $q$, since $q$ is no longer an independent variable: $q = 1 - p$. So in the previous case, we lose two degrees of freedom—one for estimating $p$ and one for independent categories. The general rule of thumb in using chi-square analysis to test for data fit to Hardy-Weinberg proportions is that the number of degrees of freedom must equal the number of phenotypes minus the number of alleles (in this case, $3 - 2 = 1$).

The chi-square analysis in table 19.2 may seem paradoxical. Because the observed allelic frequencies calculated from the original genotypic data are used to calculate the expected genotypic frequencies, it may appear to some individuals that the analysis must, by its very nature, show that the population is in Hardy-Weinberg proportions. To demonstrate that this is not necessarily the case, a counterexample appears in table 19.3. We use data similar to the Ohio sample, except that the original number of heterozygotes has been distributed equally among the two homozygote classes. The same allelic frequencies are maintained, yet the genotypic distribution differs. The chi-square value of 200.00 for these data demonstrates that the population represented in table 19.3 is not in Hardy-Weinberg proportions. Thus, a chi-square analysis of fit to the Hardy-Weinberg proportions by no means represents circular reasoning.

**Table 19.2** **Chi-Square Test of Goodness-of-Fit to the Hardy-Weinberg Proportions of a Sample of 200 Persons for MN Blood Types for Which $p = 0.76$ and $q = 0.24$**

| | MM | MN | NN | Total |
|---|---|---|---|---|
| Observed Numbers | 114 | 76 | 10 | 200 |
| Expected Proportions | $p^2$ | $2pq$ | $q^2$ | 1.0 |
| | (0.5776) | (0.3648) | (0.0576) | 1.0 |
| Expected Numbers | 115.52 | 72.96 | 11.52 | 200.0 |
| $\chi^2 = (O - E)^2/E$ | 0.020 | 0.127 | 0.201 | 0.348 |

**Table 19.3** Chi-Square Test of Goodness-of-Fit to the Hardy-Weinberg Proportions of a Second Sample of 200 Persons for MN Blood Types for Which $p = 0.76$ and $q = 0.24$ and Heterozygotes Are Absent

| | MM | MN | NN | Total |
|---|---|---|---|---|
| Observed Numbers | 152 | 0 | 48 | 200 |
| Expected Proportions | $p^2$ | $2pq$ | $q^2$ | 1.0 |
| | (0.5776) | (0.3648) | (0.0576) | 1.0 |
| Expected Numbers | 115.52 | 72.96 | 11.52 | 200.0 |
| $\chi^2 = (O - E)^2/E$ | 11.52 | 72.96 | 115.52 | 200.00 |

### *PKU*

Circumstances sometimes do not allow us to test for Hardy-Weinberg proportions. In the case of a dominant trait, for example, allelic frequencies cannot be calculated from the genotypic classes because the homozygous dominant individuals cannot be distinguished from the heterozygotes. However, we can estimate allelic frequencies by assuming that the Hardy-Weinberg equilibrium exists and, thereby, assuming that the frequency of the recessive homozygote is $q^2$, from which $q$ and then $p$ can be estimated.

If, for example, Hardy-Weinberg equilibrium is assumed for a disease such as phenylketonuria (PKU), which is expressed only in the homozygous recessive state, it is possible to calculate the proportion of the population that is heterozygous (carriers of the PKU allele). But is it fair to assume Hardy-Weinberg equilibrium here? Until recent medical advances allowed intervention, there was a good deal of selection against individuals with PKU, who were usually mentally retarded. Thus the assumption of no selection, required for equilibrium, is violated. However, only one child in ten thousand live births has PKU. When a genotype is as rare as one in ten thousand, selection has a negligible effect on allelic frequencies. Therefore, because of the rarity of the trait, we can assume Hardy-Weinberg equilibrium and calculate

$$\text{frequency of recessive homozygote} = q^2 = 1/10{,}000 = 0.0001$$

so,

$$q = \sqrt{0.0001} = 0.01$$

and

$$p = 1 - q = 0.99$$

Therefore,

$$\text{frequency of normal homozygote} = p^2 = (0.99)^2 \cong 0.98 \text{ or } 98 \text{ in } 100$$

$$\text{frequency of heterozygote} = 2pq = 2(0.01)(0.99) \cong 0.02 \text{ or } 2 \text{ in } 100.$$

By assuming the Hardy-Weinberg equilibrium, we have discovered something not intuitively obvious: A recessive gene causing a trait as rare as one in ten thousand is carried in the heterozygous state by one individual in fifty. Obviously, the chi-square test cannot be used to verify the Hardy-Weinberg proportions since we derived the allelic frequencies by assuming Hardy-Weinberg proportions to begin with. In statistical terms, the number of phenotypes minus the number of alleles $= 2 - 2 = 0$ degrees of freedom, which precludes doing a chi-square test.

## EXTENSIONS OF HARDY-WEINBERG EQUILIBRIUM

The Hardy-Weinberg equilibrium can be extended to include, among other cases, multiple alleles and multiple loci.

### *Multiple Alleles*

#### *Multinomial Expansion*

The expected genotypic array under Hardy-Weinberg equilibrium is $p^2$, $2pq$, and $q^2$, which form the terms of the binomial expansion $(p + q)^2$. If males and females each have the same two alleles in the proportions of $p$ and $q$, then genotypes will be distributed as a binomial expansion in the frequencies $p^2$, $2pq$, and $q^2$ (see fig. 19.1). To generalize to more than two alleles, one need only add terms to the binomial expansion and thus create a multinomial expansion. For example, with alleles $a$, $b$, and $c$ with frequencies $p$, $q$, and $r$, the genotypic distribution should be $(p + q + r)^2$, or

$$p^2 + q^2 + r^2 + 2pq + 2pr + 2qr$$

Homozygotes will occur with frequencies $p^2$, $q^2$, and $r^2$, and heterozygotes will occur with frequencies $2pq$, $2pr$, and $2qr$. The *ABO* blood-type locus in human beings is an interesting example because it has multiple alleles and dominance.

### ABO Blood Groups

The *ABO* locus has three alleles: $I^A$, $I^B$, and $i$, with the $I^A$ and $I^B$ alleles codominant, and both dominant to the $i$ allele. These alleles control the production of a surface antigen on red blood cells (see fig. 2.13). Table 19.4 contains blood-type data from a sample of five hundred persons from Massachusetts. Is the population in Hardy-Weinberg proportions? The answer is not apparent from the data in table 19.4 alone, since there are two possible genotypes for both the A and the B phenotypes. No estimate of the allelic frequencies is possible without making assumptions about the number of each genotype within these two phenotypic classes. Is it possible to estimate the allelic frequencies? The answer is yes, if we assume that Hardy-Weinberg equilibrium exists.

One procedure follows. Let us assume that $p = f(I^A)$, $q = f(I^B)$, and $r = f(i)$. Blood type O has the *ii* genotype; if the population is in Hardy-Weinberg proportions, this genotype should occur at a frequency of $r^2$. Thus

$$f(ii) = 231/500 = 0.462 = r^2$$

and

$$r = f(i) = \sqrt{0.462} = 0.680$$

From table 19.4, we see that blood type A plus blood type O include only the genotypes $I^AI^A$, $I^Ai$, and $ii$. If the population is in Hardy-Weinberg proportions, these together should be $(p + r)^2$, in which $p^2 = f(I^AI^A)$, $2pr = f(I^Ai)$, and $r^2 = f(ii)$:

$$(p + r)^2 = (199 + 231)/500 = 0.860$$

Then, taking the square root of each side

$$p + r = \sqrt{0.860} = 0.927$$

and

$$p = 0.927 - r = 0.927 - 0.680 = 0.247$$

The frequency of allele $I^B$, $q$, can be obtained by similar logic with blood types B and O, or simply by subtraction:

$$q = 1 - (p + r) = 1 - 0.927 = 0.073$$

Thus, the Hardy-Weinberg equilibrium can be extended to include multiple alleles and can be used to make estimates of the allelic frequencies in the ABO blood groups. With ABO, it is statistically feasible to do a chi-square test because there is one degree of freedom (number of phenotypes − number of alleles = 4 − 3 = 1). We are really testing only the AB and B categories; if we did our calculations as shown, the observed and expected values of phenotypes A and O must be equal.

**Table 19.4** ABO Blood-Type Distribution in 500 Persons from Massachusetts

| Blood Type | Genotype | Number |
|---|---|---|
| A | $I^AI^A$ or $I^Ai$ | 199 |
| B | $I^BI^B$ or $I^Bi$ | 53 |
| AB | $I^AI^B$ | 17 |
| O | $ii$ | 231 |
| Total | | 500 |

## Multiple Loci

The Hardy-Weinberg equilibrium can also be extended to consider several loci at the same time in the same population. This situation deserves mention because the whole genome is likely involved in evolutionary processes and we must, eventually, consider simultaneous allelic changes in all loci segregating alleles in an organism. (Even with a high-speed computer, simultaneous consideration of many loci is a bit far off in the future.) When two loci, *A* and *B*, on the same chromosome are in equilibrium with each other, the combinations of alleles on a chromosome in a gamete follow the product rule of probability. Consider the *A* locus with alleles *A* and *a* and the *B* locus with alleles *B* and *b*, respectively, with allelic frequencies $p_A$ and $q_A$ for *A* and *a*, respectively, and $p_B$ and $q_B$ for *B* and *b*, respectively. Given completely random circumstances, the chromosome with the *A* and *B* alleles should occur at the frequency $p_A\, p_B$. This is referred to as **linkage equilibrium.** When alleles of different loci are not in equilibrium (i.e., not randomly distributed in gametes), the condition is referred to as **linkage disequilibrium.** The approach to linkage equilibrium is gradual and is a function of the recombination distance between the two loci.

For example, let's start with a population out of equilibrium so that all chromosomes are *AB* (70%) or *ab* (30%). Then $p_A = 0.7$, $q_A = 0.3$, $p_B = 0.7$, and $q_B = 0.3$. We expect the *Ab* chromosome to occur $0.7 \times 0.3 = 0.21$, or 21% of the time. The frequency of the *Ab* chromosome is zero. Assume the map distance between the two loci is 0.1; in other words, 10% of chromatids in gametes are recombinant. Initially, we consider that each locus is in Hardy-Weinberg proportions, or the frequency of *AB*/*AB* individuals = 0.49 ($0.7 \times 0.7$); the frequency of *ab*/*ab* individuals is 0.09 ($0.3 \times 0.3$); and the frequency of *AB*/*ab* individuals is 0.42 ($2 \times 0.7 \times 0.3$).

After one generation of random mating, gametes will be as follows:

from *AB*/*AB* individuals (49%): only *AB* gametes, 49% of total

from *ab*/*ab* individuals (9%): only *ab* gametes, 9% of total

from *AB*/*ab* individuals (42%):

*AB* gametes, 18.9% of total (0.45 × 0.42)

*ab* gametes, 18.9% of total (0.45 × 0.42)

*Ab* gametes, 2.1% of total (0.05 × 0.42)

*aB* gametes, 2.1% of total (0.05 × 0.42)

(The values of 18.9% and 2.1% for the dihybrids result from the fact that since map distance is 0.1, 10% of gametes will be recombinant, split equally between the two recombinant classes—5% and 5%. Ninety percent will be parental, split equally between the two parental classes—45% and 45%. Each of these numbers must be multiplied by 0.42 because the dihybrid makes up 42% of the total number of individuals.)

Although we expect 21% of the chromosomes to be of the *Ab* type, only 2.1%, 10% of the expected, appear in the gene pool after one generation of random mating. You can see that linkage equilibrium is achieved at a rate dependent on the map distance between loci. Unlinked genes, appearing 50 map units apart, also gradually approach linkage equilibrium.

Although we will not derive these extensions here, we note two others. If the frequencies of alleles at an autosomal locus differ in the two sexes, it takes two generations of random mating to achieve equilibrium. In the first generation, the allelic frequencies in the two sexes are averaged so that each sex now has the same allelic frequencies. Genotypic frequencies then come into Hardy-Weinberg proportions in the second generation. However, if the allelic frequencies differ in the two sexes for a sex-linked locus, Hardy-Weinberg proportions are established only gradually. The reasoning is straightforward. Females, with an X chromosome from each parent, average the allelic frequencies from the previous generation. However, males, who get their X chromosomes from their mothers, have the allelic frequencies of the females in the previous generation. Hence, the allelic frequencies are not the same in the two sexes after one generation of random mating, and equilibrium is achieved slowly.

## NONRANDOM MATING

The Hardy-Weinberg equilibrium is based on the assumption of random mating. Deviations from random mating come about when phenotypic resemblance or relatedness influences mate choice. When phenotypic resemblance influences mate choice, either *assortative* or *disassortative* mating occurs, depending on whether individuals choose mates on the basis of similarity or dissimilarity, respectively. For example, in human beings, assortative mating occurs for height—short men tend to marry short women, and tall men tend to marry tall women. When relatedness influences mate choice, either *inbreeding* or *outbreeding* occurs, depending on whether mates are more or less related than two randomly chosen individuals from the population. An example of inbreeding in human beings is marriage between first cousins. Both types of nonrandom mating (assortative-disassortative mating and inbreeding-outbreeding) have the same qualitative effects on the Hardy-Weinberg equilibrium: assortative mating and inbreeding increase homozygosity without changing allelic frequencies, whereas disassortative mating and outbreeding increase heterozygosity without changing allelic frequencies.

Two differences are apparent, however, between the effects of phenotypic resemblance and relatedness on mate choice. First, assortative or disassortative mating disturbs the Hardy-Weinberg equilibrium only when the phenotype and genotype are closely related. That is, if assortative mating occurs for a nongenetic trait, then the Hardy-Weinberg equilibrium will not be distorted. Inbreeding and outbreeding affect the genome directly. A second difference between the two types of mating is that the effects of inbreeding or outbreeding are felt across the whole genome, whereas the disturbances to the equilibrium caused by assortative and disassortative mating occur only for the particular trait being considered (and for closely linked loci). Given the similarities in the consequences of the two types of matings, we will concentrate our discussion on inbreeding.

### Inbreeding

Inbreeding comes about in two ways: (1) the systematic choice of relatives as mates and (2) the subdivision of a population into small subunits, leaving individuals little choice but to mate with relatives. We will concentrate on inbreeding as the systematic choice of relatives as mates. The consequences of both are similar.

#### Common Ancestry

An inbred individual is one whose parents are related—that is, there is **common ancestry** in the family tree. The extent of inbreeding thus depends on the degree of common ancestry that the parents of an inbred individual share. When mates share ancestral genes, each may pass on copies of the same ancestral allele to their offspring. An inbred individual can then carry identical copies of a single ancestral allele. In other words, an individual of *aa* genotype is homozygous and, if it is possible that the *a* allele from each parent is a length of DNA originally copied from a common ancestor, the *aa* individual is said to be inbred.

The first observable effect of inbreeding is the expression of hidden recessives. In human beings, each individual carries, on the average, about four **lethal-equivalent alleles,** alleles that kill when paired to form a homozygous genotype (box 19.1). In many, and probably most, human societies, zygotes are generally heterozygous for these lethal alleles because of a cultural pattern of outbreeding, mating with nonrelatives. Rarely does an outbred zygote receive the same recessive lethal from each parent. Dominance acts to mask the expression of deleterious recessive alleles. But, in the process of inbreeding, when the zygote may receive copies of the same ancestral allele from each parent, there is a substantial increase in the probability that a deleterious allele will pair to form a homozygous genotype (fig. 19.2). Inbreeding can result in spontaneous abortions (miscarriages), fetal deaths, and congenital deformities. In many species, however, inbreeding—even self-fertilization—occurs normally. These species usually do not have the problem with lethal equivalents that species that normally outbreed do. Through time, species that normally inbreed have had these deleterious alleles mostly eliminated, presumably by natural selection. Inbreeding has even been used successfully for artificial selection in livestock and crop plants.

From our previous discussion, you can see that there are two types of homozygosity—**allozygosity,** in which two alleles are alike but unrelated (not copies of the same ancestral allele) and **autozygosity,** in which two alleles have **identity by descent** (i.e., are copies of the same ancestral allele). An **inbreeding coefficient, *F*,** can be defined as the probability of autozygosity, or the probability that the two alleles in an individual at a given locus are identical by descent. This coefficient can range from zero, at which point there is no inbreeding, to one, at which point it is certain an individual is autozygous.

### Increased Homozygosity from Inbreeding

What are the effects of inbreeding on the Hardy-Weinberg equilibrium? Let us for a moment return to the gene pool concept to produce zygotes. Assume that an allele drawn from this gene pool is of the *A* type, drawn with a probability of $p$. On the second draw, the probability of autozygosity, that is, of drawing a copy of the same allele *A*, is $F$, the inbreeding coefficient. Thus the probability of an autozygous *AA* individual is $pF$. On the second draw, however, with probability $(1 - F)$, either the *A* or *a* allele can be drawn, with probabilities of $p^2(1 - F)$ and $pq(1 - F)$, respectively. Note that a second *A* allele produces a homozygote that is not inbred (allozygous). If the first allele drawn was an *a* allele, with probability $q$, then the probability of drawing the same allele (copy of the same ancestral allele) is $F$, and thus the probability of autozygosity is $qF$. However, the probability of drawing an *a* or *A* allele that does not contribute to inbreeding is $(1 - F)$ and, therefore, the probability of an *aa* or *Aa* genotype is $q^2(1 - F)$ and $pq(1 - F)$, respectively. These calculations are summarized in table 19.5, a summary of the genotypic proportions in a population with inbreeding.

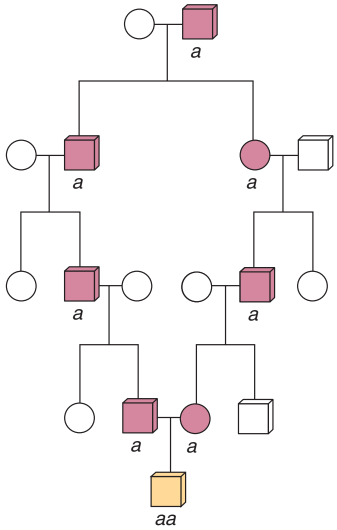


**Figure 19.2** Homozygosity by descent of copies of the same ancestral allele, *a*. The individual at the bottom of the pedigree is inbred with the *aa* genotype.

Several points emerge from table 19.5. First, when the inbreeding coefficient is zero (completely random mating), the table reduces to Hardy-Weinberg proportions. Second, compared with Hardy-Weinberg proportions, inbreeding increases the proportion of homozygotes in the population (identity by descent implies homozygosity). With complete inbreeding ($F = 1$), only homozygotes will occur in the population.

How does inbreeding affect allelic frequencies? Recall that an allelic frequency is calculated as the frequency of homozygotes for one allele plus half the frequency of the heterozygotes. Here we let $p_{n+1}$ be the frequency of the *A* allele after one generation of inbreeding:

$$\begin{aligned}
p_{n+1} &= p^2(1 - F) + pF + (1/2)(2pq)(1 - F) \\
&= p^2(1 - F) + pF + pq(1 - F) \\
&= p^2 + pq + F(p - p^2 - pq) \\
&= p(p + q) + pF(1 - p - q) \\
&= p(1) + pF(0) \\
&= p
\end{aligned}$$

**BOX 19.1**

## Experimental Methods

### *The Determination of Lethal Equivalents*

The average person carries about four lethal-equivalent alleles that are hidden because they are recessive. Four lethal equivalents means four alleles that are lethal when homozygous, or eight alleles conferring a 50% chance of mortality when homozygous, or any similar combination of lethal and semilethal alleles. The exact arrangement cannot be determined with current analytical methods. We arrive at the estimate of hidden defective and lethal alleles by using inbreeding data.

J. Crow and M. Kimura, in 1970, analyzed data showing that in Swedish families in which marriages occurred between first cousins, between 16 and 28% of the offspring had genetic diseases. For unrelated parents, the comparable figure is between 4 and 6%. Therefore, it is estimated that the offspring of first cousins have an added risk of 12 to 22% of having a genetic defect. The children of first cousins have an inbreeding coefficient of one-sixteenth. Hence, a theoretical individual who is completely inbred has the risk of genetic defect increased sixteenfold over an individual whose parents are first cousins. If 100% risk is considered 1 lethal equivalent, then a completely inbred individual would carry 2 to 3.5 lethal equivalents (16 × 12%–16 × 22%). However, a completely inbred individual is, in essence, a doubled gamete. Since our interest is in the number of deleterious alleles a normal person carries, it is necessary to further multiply the risk by a factor of two to determine the number of lethal-equivalent alleles carried by a normal individual. The conclusion is that the average person carries the equivalent of four to seven alleles that would, in the homozygous state, cause a genetic defect.

A similar calculation can be made using viability data rather than genetic defects to determine the occurrence of lethal equivalents. A study from rural France, also analyzed by Crow and Kimura, showed that the mortality rate of offspring of first cousins was 25%, whereas the analogous figure for the offspring of unrelated parents was about 12%, an increased risk of 13% for the offspring of cousins. Multiplying this risk figure of 0.13 by 32 (16 × 2) presents a figure of four lethal equivalents per average person in the population. In 1971, L. Cavalli-Sforza and W. Bodmer, using data primarily from Japanese populations, reported an estimate of about two lethal equivalents per average person. Despite some interpopulation differences in these estimates, they are about the same order of magnitude—two to seven lethal equivalents per person.

Thus, inbreeding does not change allelic frequencies. We can also see intuitively that inbreeding affects zygotic combinations (genotypes), but not allelic frequencies: Although inbreeding may determine the genotypes of offspring, inbreeding does not change the numbers of each allele that an individual transmits into the next generation.

In summary, inbreeding causes an increase in homozygosity, affects all loci in a population equally, and, in itself, has no effect on allelic frequencies, although it can expose deleterious alleles to selection. The results of inbreeding are evident in the appearance of recessive traits that are often deleterious. Inbreeding increases the rate of fetal deaths and congenital malformations in human beings and in other species that normally outbreed. In outbred agricultural crops and farm animals, decreases in size, fertility, vigor, and yield often result from inbreeding. Once deleterious traits appear due to inbreeding, natural selection can cause their removal from the population. However, in species adapted to inbreeding, including many crop plants and farm animals, inbreeding does not expose deleterious alleles because those alleles have generally been eliminated already.

## *Pedigree Analysis*

### *Path Diagram Construction*

The inbreeding coefficient, *F*, of an individual (the probability of autozygosity) can be determined by pedigree analysis. This is done by converting a pedigree to a **path diagram** by eliminating all extraneous individuals, those who cannot contribute to the inbreeding coefficient of the individual in question. A path diagram shows the direct line of descent from common ancestors. An example of the conversion of a pedigree to a path diagram is shown in figure 19.3, in which individuals C and F are omitted from the path of descent because they are not related to anyone on the other side of the family tree and, therefore, do not contribute to the "common ancestry" of individual I. The pedigree in figure 19.3 shows an offspring who is the daughter of first cousins. Since first

**Table 19.5** Genotypic Proportions in a Population with Inbreeding

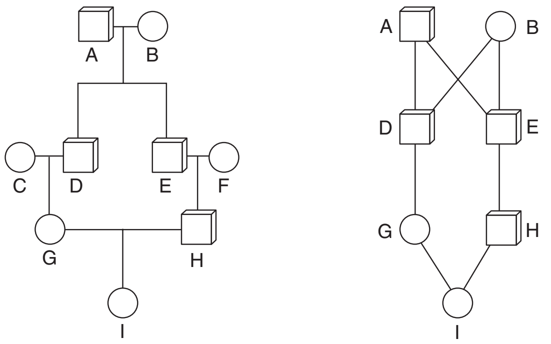
| Genotype | Due to Random Mating $(1-F)$ | | Due to Inbreeding $(F)$ | | Observed Proportions |
|---|---|---|---|---|---|
| *AA* | $p^2(1-F)$ | + | $p$F | = | $p^2 + Fpq$ |
| *Aa* | $2pq(1-F)$ | | | = | $2pq(1-F)$ |
| *aa* | $q^2(1-F)$ | + | $q$F | = | $q^2 + Fpq$ |
| Total | $(p^2 + 2pq + q^2)(1-F)$ | + | $(p+q)F$ | = | |
| | $(1-F)$ | + | $F$ | = | 1 |

cousins are the offspring of siblings, they share a set of common grandparents. Thus, individual I can be autozygous for alleles from either ancestor A or B, her great-grandparents. The path diagram shows the only routes by which autozygosity can occur.

The inbreeding coefficient of the offspring of first cousins can be calculated as follows. The path diagram of figure 19.3 is shown again in figure 19.4, with lowercase letters designating gametes. Two paths of autozygosity appear in this diagram, one path for each grandparent as a common ancestor: A to D and E, then to G and H, and finally to I; or B to D and E, then to G and H, and finally to I.

In the path with A as the common ancestor, A contributes a gamete to D and a gamete to E. The probability is one-half that D and E each carry a copy of the same allele. That is, there are four possible allelic combinations for the two gametes, $a_1$ and $a_2$: *A-A*; *A-a*; *a-A*; and *a-a*. Of these combinations, the first and last (*A-A* and *a-a*) give a copy of the same allele to the two offspring, D and E, and can thus contribute to autozygosity. The probability that gametes $a_1$ and $d$ carry copies of the same allele is one-half, and the probability that $d$ and $g$ carry copies of the same allele is also one-half. Similarly, on the other side of the pedigree, the probability is one-half that $a_2$ and $e$ carry copies of the same allele and one-half that $e$ and $h$ carry copies of the same allele. Thus, the overall probability that the alleles that $g$ and $h$ carry are identical by descent (autozygous) is $(1/2)^5$. In general, it would be $(1/2)^n$ for each path, where $n$ is the number of ancestors in the path.

You may have spotted an additional factor here. Of the possible combinations of allelic copies passed on to D and E, one-half (*A-A* and *a-a*) are autozygous combinations. However, the other half of the combinations, *A-a* and *a-A*, can lead to autozygosity if A is itself inbred. If we let $F_A$ be the inbreeding coefficient of A (the probability that any two alleles at a locus in A are identical by descent), then $F_A$ is the probability that the *A-a* and *a-A* combinations are also autozygous. Thus, the probability that a common ancestor, A, passes on copies of an



Pedigree Path diagram

**Figure 19.3** Conversion of a pedigree to a path diagram. This pedigree depicts the mating of first cousins. In the path diagram, all extraneous individuals are removed, leaving only those who could contribute to the inbreeding of individual I. Individuals in the line of descent are connected directly with *straight lines,* indicating the paths along which gametes are passed.

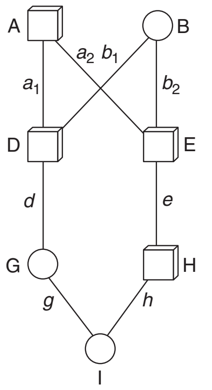


**Figure 19.4** The path diagram of the mating of first cousins with gametes labeled in *lowercase letters.*

identical ancestral allele is $1/2 + (1/2)F_A$, or $(1/2)(1 + F_A)$. In other words, there is a one-half probability that the alleles transmitted from A to D and E are copies of the same allele. In the other half of the cases, these alleles can be identical if A is inbred. The probability of identity of A's two alleles is $F_A$. The expression for the inbreeding coefficient of I, $F_I$, can now be changed from $(1/2)^n$ by substituting $(1/2)(1 + F_A)$ for one of the (1/2)s to

$$F_I = (1/2)^n(1 + F_A)$$

This equation accounts only for the inbreeding of I by the path involving the common ancestor, A, and does not account for the symmetrical path with B as the common ancestor. To obtain the total probability of inbreeding, the values from each path must be added (because these are mutually exclusive events; see chapter 4). Thus the complete formula for the inbreeding coefficient of the offspring of first cousins is

$$F_I = \Sigma[(1/2)^n(1 + F_J)] \qquad \textbf{(19.1)}$$

in which $F_I$ is the probability that the two alleles in I are identical by descent, $n$ is the number of ancestors in a given path, $F_J$ is the inbreeding coefficient of the common ancestor of that path, and all paths are summed.

In the example of the mating of first cousins (fig. 19.4)

$$F_I = (1/2)^5(1 + F_A) + (1/2)^5(1 + F_B)$$

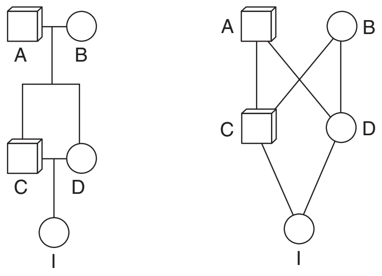
If we assume that $F_A$ and $F_B$ are zero (which we must assume when the pedigrees of A and B are unknown), then

$$F_I = 2(1/2)^5 = (1/2)^4 = 0.0625$$

This can be interpreted to mean that about 6.25% of individual I's loci are autozygous, or that there is a 6.25% chance of autozygosity at any one of I's loci.

The inbreeding coefficient of the offspring of siblings (fig. 19.5) can also be calculated, assuming that A and B are not themselves inbred ($F_A$ and $F_B$ are zero), as

$$F_I = 2(1/2)^3 = 0.25$$



Pedigree Path diagram

**Figure 19.5** Conversion of a sib-mating pedigree to a path diagram. Individual I is inbred.

Thus, about 25% of the loci in an offspring of siblings are autozygous.

### *Path Diagram Rules*

The following points should be kept in mind when calculating an inbreeding coefficient:

1. All possible paths must be counted. A path is possible if gametes can actually pass in that direction. Paths that violate the rules of inheritance cannot be used. For example, in figure 19.4, the following path is unacceptable: I G E A D H I.
2. In any path, an individual can be counted only once.
3. Every path must have one and only one common ancestor. The inbreeding coefficient of any other individual in the path is immaterial.

In figure 19.6, we present a complex pedigree produced from repeated sib mating, a pattern found in livestock and laboratory animals. This pedigree has two interesting points. First, common ancestors occur in several different generations. Second, some of the paths are complex. Thus, we must be sure to count all paths (paths 5 and 6 might not be immediately obvious). Although not shown in figure 19.6, one of the common ancestors, A, is also inbred ($F_A = 0.05$)—a fact that we must take into consideration in paths 3 and 5. Thus, $F_I$ is as follows:

From path 1: $(1/2)^3$ $= 0.1250$

From path 2: $(1/2)^3$ $= 0.1250$

From path 3: $(1/2)^5(1 + 0.05)$ $= 0.0328$

From path 4: $(1/2)^5$ $= 0.0313$

From path 5: $(1/2)^5(1 + 0.05)$ $= 0.0328$

From path 6: $(1/2)^5$ $= 0.0313$

---

$F_I = 0.3782$

## Population Analysis

It is also possible to define the inbreeding coefficient, $F$, of a population as the relative reduction in heterozygosity in the population due to inbreeding. In an individual, $F$ is the probability of autozygosity; it represents an increase in homozygosity, which is therefore a decrease in heterozygosity. In a population, it also represents the reduction in heterozygosity. From the definition, we can calculate the population $F$ as follows:

$$F = \frac{(2pq - H)}{2pq}$$

where $H$ is the actual proportion of heterozygotes in a population, and $2pq$ is the expected proportion of heterozygotes based on Hardy-Weinberg proportions. This equation reduces to

$$F = 1 - \frac{H}{2pq} \tag{19.2}$$

This equation shows that when $H = 2pq$, $F$ is zero, meaning that there is no decrease in heterozygotes and therefore, apparently, no inbreeding. When there are no heterozygotes, $F = 1$. This could be the case in a completely inbred population—for example, a self-fertilizing plant species.

As an example of an intermediate case, take the sample of one hundred individuals segregating the $A_1$ and $A_2$ alleles at the $A$ locus: $A_1A_1$, fifty-four; $A_1A_2$, thirty-two; and $A_2A_2$, fourteen. In this example, $p = 0.7$, $q = 0.3$, and $H = 0.32$. Since $2pq = 0.42$, $H/2pq = 0.32/0.42 = 0.76$, and $F = 1 - 0.76$, or 0.24. Thus, the inbreeding coefficient of this population is 0.24; there is a 24% reduction in heterozygotes, due presumably to inbreeding.

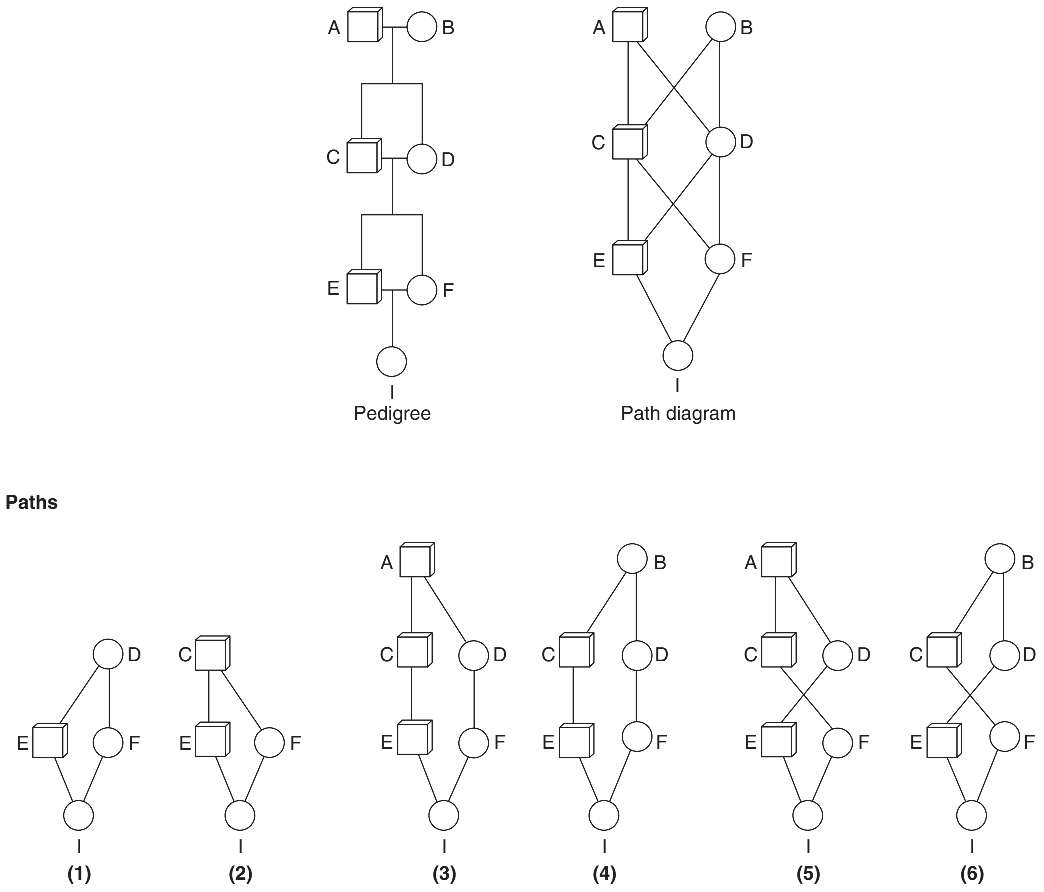


**Figure 19.6** Pedigree and path diagram of two generations of sib matings. The six paths involving the potential for autozygosity are shown. $F_A = 0.05$. The paths involve common ancestors in two generations.

## SUMMARY

**STUDY OBJECTIVE 1:** To understand the concept of population-level genetic processes 553–554

In a large, randomly mating population of sexually reproducing diploid organisms, not subject to the influences of mutation, migration, or selection, an equilibrium will be achieved for an autosomal locus with two alleles.

**STUDY OBJECTIVE 2:** To learn the assumptions and nature of the Hardy-Weinberg equilibrium and its extensions 554–557

The Hardy-Weinberg equilibrium predicts that (1) allelic frequencies ($p$, $q$) will not change from generation to generation; (2) genotypes will occur according to the binomial distribution $p^2 = f(AA)$, $2pq = f(Aa)$, and $q^2 = f(aa)$; and (3) if perturbed, equilibrium will reestablish itself in just one generation of random mating.

**STUDY OBJECTIVE 3:** To test whether a population is in Hardy-Weinberg equilibrium 557–560

To determine whether a population is in Hardy-Weinberg proportions, the observed and expected distribution of genotypes can be compared by the chi-square statistical test. In some circumstances, when it is reasonable to assume equilibrium, we can estimate allelic and genotypic frequencies even when dominance occurs. The Hardy-Weinberg equilibrium is easily extended to the prediction of the frequencies of multiple alleles, multiple loci, and different frequencies of alleles in the two sexes, for both sex-linked and autosomal loci.

**STUDY OBJECTIVE 4:** To analyze the process and consequences of nonrandom mating in diploid populations 560–565

Random mating is required for the Hardy-Weinberg equilibrium to hold. Deviations from random mating fall into two categories, depending on whether phenotypic resemblance or relatedness is involved in mate choice. Phenotypic resemblance is the basis for assortative and disassortative mating, in which individuals choose similar or dissimilar mates, respectively. Assortative mating causes increased homozygosity only among loci controlling the traits that influence mate choice. There are no changes in allelic frequencies. Similarly, disassortative mating causes increased heterozygosity without changing allelic frequencies.

Mating among relatives, or inbreeding, is represented by $F$, the inbreeding coefficient, which measures the probability of autozygosity (homozygosity by descent). It can be calculated from pedigrees by using the formula

$$F = \Sigma[(1/2)]^n(1 + F_J)]$$

where $n$ is the number of ancestors in a given path and $F_J$ is the inbreeding coefficient of the common ancestor of that path. Inbreeding exposes recessive deleterious traits already present in the population and causes homozygosity throughout the genome. It does not, by itself, change allelic frequencies. $F$ can also be calculated from the reduction in heterozygosity in a population.

## SOLVED PROBLEMS

**PROBLEM 1:** One hundred fruit flies (*Drosophila melanogaster*) from California were tested for their genotype at the alcohol dehydrogenase locus using starch-gel electrophoresis. Two alleles were present, *S* and *F*, for slow and fast migration, respectively. The following results were noted: *SS*, sixty-six; *SF*, twenty; *FF*, fourteen. What are the allelic and genotypic frequencies in this population?

***Answer:*** Since the sample size is one hundred, the proportions of the three genotypes, *SS*, *SF*, and *FF*, are 0.66, 0.20, and 0.14, respectively. We can calculate allelic frequencies directly from these genotypes, remembering that the frequency of an allele is the frequency of its homozygote plus half the frequency of the heterozygote, or

$$\begin{aligned} p &= f(S) = f(SS) + (1/2)f(SF) \\ &= 0.66 + (1/2)(0.20) = 0.76 \end{aligned}$$

$$\begin{aligned} q &= f(F) = f(FF) + (1/2)f(SF) \\ &= 0.14 + (1/2)(0.20) = 0.24 \end{aligned}$$

Alternatively, we could get allelic frequencies by counting alleles. Thus,

$$p = \frac{2 \times \text{number of } SS + \text{number of } SF}{2 \times \text{total number}} = \frac{2(66) + 20}{2(100)} = \frac{152}{200} = 0.76$$

$$p = \frac{2 \times \text{number of } FF + \text{number of } SF}{2 \times \text{total number}} = \frac{2(14) + 20}{2(100)} = \frac{48}{200} = 0.24$$

**PROBLEM 2:** Is the population described in problem 1 in Hardy-Weinberg equilibrium?

***Answer:*** We can determine whether the numbers of the three genotypes (*SS*, *SF*, and *FF*) are in Hardy-Weinberg proportions through the chi-square statistical test. The observed numbers of the three genotypes are sixty-six, twenty, and fourteen, respectively. Using allelic frequencies of $p = f(S) = 0.76$ and $q = f(F) = 0.24$, we expect $p^2$, $2pq$, and $q^2$, respectively, of the three genotypes. That is,

$$p^2 = (0.76)^2 = 0.5776, \text{ or } 57.76 \text{ in } 100$$

$$2pq = 2(0.76)(0.24) = 0.3648, \text{ or } 36.48 \text{ in } 100$$

$$q^2 = (0.24)^2 = 0.0576, \text{ or } 5.76 \text{ in } 100$$

We can now set up a chi-square table as follows:

| | *SS* | *SF* | *FF* | Total |
|---|---|---|---|---|
| Observed Numbers | 66 | 20 | 14 | 100 |
| Expected Proportions | $p^2$ | $2pq$ | $q^2$ | 1.0 |
| | (0.5776) | (0.3648) | (0.0576) | 1.0 |
| Expected Numbers | 57.76 | 36.48 | 5.76 | 100 |
| $\chi^2 = (O - E)^2/E$ | 1.176 | 7.445 | 11.788 | 20.408 |

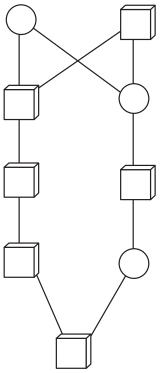
The critical chi-square value (0.05 at one degree of freedom) is 3.841, so we reject the hypothesis that this population is in Hardy-Weinberg proportions. From inspection of the table, it appears that there are too few heterozygotes and too many homozygotes, indicating that inbreeding could be the cause of the discrepancy.

**PROBLEM 3:** Convert the pedigree in figure 19.2 into a path diagram, and determine the inbreeding coefficient of the inbred individual, assuming that the common ancestors are not themselves inbred.

***Answer:*** There are two paths (see the figure), each with seven ancestors. Thus, the inbreeding coefficient is

$$F = \Sigma\,[(1/2)^n\,(1 + F_J)] = 2(1/2)^7 = 0.016$$

Hence, the inbreeding coefficient is 0.016; about 1.6% of the loci of the inbred individual are autozygous.



# EXERCISES AND PROBLEMS*

## HARDY-WEINBERG EQUILIBRIUM

1. One hundred persons from a small town in Pennsylvania were tested for their MN blood types. Is the population they represent in Hardy-Weinberg proportions? The genotypic data are: *MM*, forty-one; *MN*, thirty-eight; and *NN*, twenty-one.
2. From the following two sets of data, calculate allelic and genotypic frequencies, and determine whether the populations are in Hardy-Weinberg proportions. Do a statistical test if one is appropriate.
   a. Allele *A* is dominant to *a*; *A*-, 91; *aa*, 9.
   b. Electrophoretic alleles *F* and *S* are codominant at the malate dehydrogenase locus in *Drosophila; FF*, 137; *FS*, 196; *SS*, 87.
3. The dominant ability to taste PTC comes from the allele *T*. Among a sample of 215 individuals from a population in Vancouver, 150 could detect the taste of PTC, and 65 could not. Calculate the allelic frequencies of *T* and *t*. Is the population in Hardy-Weinberg proportions?

---

* Answers to selected exercises and problems are on page A-21.

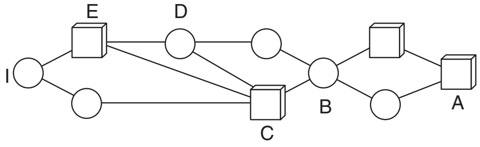
4. The frequency of children homozygous for the recessive allele for cystic fibrosis is about one in twenty-five hundred. What is the percentage of heterozygotes in the population?
5. PTC tasting is dominant in human beings.
   - **a.** Should most human populations be heading toward a 3:1 ratio of tasters to nontasters? Explain.
   - **b.** Confronted with a population sample of human beings of unknown origin, would you expect more or less than half the sample to be tasters?
6. Graph the relationship of the proportions of genotype (*AA, Aa, aa*) as allelic frequencies change.
7. A particular recessive disorder is present in one in ten thousand individuals. If the population is in Hardy-Weinberg equilibrium, what are the frequencies of the two alleles?
8. What allelic frequency will generate twice as many recessive homozygotes as heterozygotes?
9. Assume brown eye color is the result of a dominant allele at one locus. Attack or defend mathematically the following statement: With time, the frequency of brown-eyed individuals will increase, until about three out of four individuals are brown-eyed.
10. A particular human population has five hundred *MM* individuals, three hundred *MN*, and seven hundred *NN*. Calculate the allelic frequencies, and determine whether the population is in Hardy-Weinberg equilibrium.
11. Assume random mating occurs among the individuals of the population described in problem 10. What will be the frequency of each type of individual in the next generation?
12. On a small island, 235 mating individuals are all true-breeding for brown eyes. An epidemic eliminates all the population except ten young women, two young men, and four older (postmenopausal) women. A boatload of foreigners arrives; the foreign population consists of six heterozygous brown-eyed females, four homozygous brown-eyed males, and ten blue-eyed males. Assuming that one locus controls eye color, that mating is random with respect to eye color, and that each male and female capable of breeding does so, calculate the genotypic frequencies of their offspring.
13. In a given population, only the $I^A$ and $I^B$ alleles are present in the ABO system; there are no individuals with type O blood or with $i$ alleles. If two hundred people have type A blood, seventy-five have type AB blood, and twenty-five have type B blood, what are the allelic frequencies in this population?

**EXTENSIONS OF HARDY-WEINBERG EQUILIBRIUM**

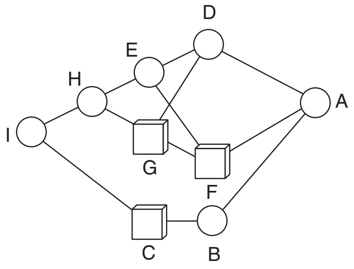
14. The following data are *ABO* phenotypes from a population sample of one hundred persons. Determine the frequencies of the three alleles: type A, seven; type B, seventy-two; type AB, twelve; type O, nine. What do you have to assume? Is the population in Hardy-Weinberg proportions?
15. How quickly and in what manner is Hardy-Weinberg equilibrium achieved under the following initial conditions (assuming a diploid, sexually reproducing population)?
   - **a.** One locus, five alleles
   - **b.** Two unlinked loci, two alleles each
16. A sample of fruit flies was testcrossed to determine the allelic arrangements of two linked loci in the gametes of that generation. With the following data, can you determine whether linkage equilibrium holds? Gametic arrangements are *AB*, fifty-eight; *ab*, eight; *Ab*, twelve; and *aB*, twenty-two.
17. In a large, randomly mating human population, the frequencies of the $I^A$, $I^B$, and $i$ alleles are 0.7, 0.2, and 0.1, respectively. Calculate the expected frequencies for each blood type.
18. In a human population of one hundred people, seventeen have type A blood, seventeen have type B, two have type AB, and sixty-four have type O. If this population is in equilibrium, what are the allelic frequencies?

**NONRANDOM MATING**

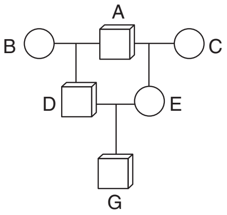
19. Under what circumstances is inbreeding deleterious?
20. What is the inbreeding coefficient of I in the following pedigree? Assume that the inbreeding coefficients of other members of the pedigree are zero unless other information tells you differently.



21. What is the inbreeding coefficient of individual I in this pedigree? $F_A = 0.01$; $F_B = 0.02$; $F_C = 0.02$.



**22.** The following is the pedigree of an offspring produced by the mating of half siblings. Individuals A and C have inbreeding coefficients of 0.2; all others are zero. Convert the pedigree to a path diagram and determine the inbreeding coefficient of individual G.



**23.** Given the population in Exercises and Problems problem 1, what is its inbreeding coefficient?

**24.** In a sample of one hundred people, are fourteen *MM*, thirty-two *MN*, and fifty-four *NN* individuals. Calculate the inbreeding coefficient.

**25.** If, in a population with two alleles at an autosomal locus, $p = 0.8$, $q = 0.2$, and the frequency of heterozygotes is 0.20, what is the inbreeding coefficient?

## CRITICAL THINKING QUESTIONS

**1.** Prove that two generations are needed for the establishment of Hardy-Weinberg proportions when an autosomal locus with two alleles in a sexually reproducing species has frequencies of the two alleles that differ in the two sexes.

**2.** What might the ramifications to conservation efforts be of zoos maintaining captive breeding programs for rare and endangered species?

*Suggested Readings for chapter 19 are on page B-19.*

# 20

# POPULATION GENETICS

## *Processes That Change Allelic Frequencies*

### STUDY OBJECTIVES

1. To develop ways to analyze population genetics problems 571
2. To analyze the effects of mutation, migration, and population size on the Hardy-Weinberg equilibrium 571
3. To study the ways in which natural selection results in organisms adapted to their environments 577

### STUDY OUTLINE

**Models for Population Genetics** 571
**Mutation** 571
    Mutational Equilibrium 571
    Stability of Mutational Equilibrium 571
**Migration** 573
**Small Population Size** 574
    Sampling Error 574
    Simulation of Random Genetic Drift 575
    Founder Effects and Bottlenecks 576
**Natural Selection** 577
    How Natural Selection Acts 577
    Selection Against the Recessive Homozygote 578
    Selection-Mutation Equilibrium 581
    Types of Selection Models 581
*Summary* 585
*Solved Problems* 585
*Exercises and Problems* 586
*Critical Thinking Questions* 587
**Box 20.1 A General Computer Program to Simulate the Approach to Allelic Equilibrium Under Heterozygous Advantage** 583

Natural selection works on the variation found in nature, shown here by different banding patterns in tree snails (*Liguus fasciatus*), found mainly in southern Florida. (© J. H. Robinson/Photo Researchers, Inc.)

We continue our discussion of the genetics of the evolutionary process. This chapter is devoted to a discussion of some of the effects of violating, or relaxing, the assumptions of the Hardy-Weinberg equilibrium other than random mating, which we discussed in chapter 19. Here we consider the effects of mutation, migration, small population size, and natural selection on the Hardy-Weinberg equilibrium. These processes usually change allelic frequencies.

## MODELS FOR POPULATION GENETICS

The steps we need to take to solve for equilibrium in population genetics models follow the same general pattern regardless of what model we are analyzing. We emphasize that these models were developed to help us understand the genetic changes taking place in a population. The models shed light on nonintuitive processes and help quantify intuitive processes. The steps in the models can be outlined as follows:

1. Set up an algebraic model.
2. Calculate allelic frequency in the next generation, $q_{n+1}$.
3. Calculate change in allelic frequency between generations, $\Delta q$.
4. Calculate the equilibrium condition, $\hat{q}$ ($q$-hat), at $\Delta q = 0$.
5. Determine, when feasible, if the equilibrium is stable.

## MUTATION

### Mutational Equilibrium

Mutation affects the Hardy-Weinberg equilibrium by changing one allele to another and thus changing allelic and genotypic frequencies. Consider a simple model in which two alleles, $A$ and $a$, exist. $A$ mutates to $a$ at a rate of $\mu$ (mu), and $a$ mutates back to $A$ at a rate of $\nu$ (nu):

$$A \underset{\nu}{\overset{\mu}{\rightleftarrows}} a$$

If $p_n$ is the frequency of $A$ in generation $n$ and $q_n$ is the frequency of $a$ in generation $n$, then the new frequency of $a$, $q_{n+1}$, is the old frequency of $a$ plus the addition of $a$ alleles from forward mutation and the loss of $a$ alleles by back mutation. That is,

$$q_{n+1} = q_n + \mu p_n - \nu q_n \qquad \textbf{(20.1)}$$

in which $\mu p_n$ is the increment of $a$ alleles added by forward mutation, and $\nu q_n$ is the loss of $a$ alleles due to back mutation. Equation 20.1 takes into account not only the rate of forward mutation, $\mu$, but also $p_n$, the frequency of $A$ alleles available to mutate. Similarly, the loss of $a$ to $A$ alleles is the product of both the rate of back mutation, $\nu$, and the frequency of the $a$ allele, $q_n$. Equation 20.1 completes the second modeling step, derivation of an expression for $q_{n+1}$, allelic frequency after one generation of mutation pressure. The third step is to derive an expression for the change in allelic frequency between two generations. This change ($\Delta q$) is simply the difference between the allelic frequency at generation $n + 1$ and the allelic frequency at generation $n$. Thus, for the $a$ allele

$$\Delta q = q_{n+1} - q_n = (q_n + \mu p_n - \nu q_n) - q_n \qquad \textbf{(20.2)}$$

which simplifies to

$$\Delta q = \mu p_n - \nu q_n \qquad \textbf{(20.3)}$$

The next step in the model is to calculate the equilibrium condition $\hat{q}$, or the allelic frequency when there is no change in allelic frequency from one generation to the next—that is, when $\Delta q$ (equation 20.3) is equal to zero:

$$\Delta q = \mu p_n - \nu q_n = 0 \qquad \textbf{(20.4)}$$

Thus,

$$\mu p_n = \nu q_n \qquad \textbf{(20.5)}$$

Then, substituting $(1 - q_n)$ for $p_n$ (since $p = 1 - q$), gives

$$\mu(1 - q_n) = \nu q_n$$

or, by rearranging:

$$\hat{q} = \frac{\mu}{\mu + \nu} \qquad \textbf{(20.6)}$$

And, since $p + q = 1$,

$$\hat{p} = \frac{\nu}{\mu + \nu} \qquad \textbf{(20.7)}$$

We can see from equations 20.6 and 20.7 that an equilibrium of allelic frequencies does exist. Also, the equilibrium value of allele $a$ ($\hat{q}$) is directly proportional to the relative size of $\mu$, the rate of forward mutation toward $a$. If $\mu = \nu$, the equilibrium frequency of the $a$ allele ($\hat{q}$) will be 0.5. As $\mu$ gets larger, the equilibrium value shifts toward higher frequencies of the $a$ allele.

### Stability of Mutational Equilibrium

Having demonstrated that allelic frequencies can reach an equilibrium due to mutation, we can ask whether the mutational equilibrium is stable. A stable equilibrium is one that returns to the original equilibrium point after being perturbed. An unstable equilibrium is one that will not return after being perturbed but, rather, continues to move away from the equilibrium point. As we mentioned in the last chapter, the Hardy-Weinberg equilibrium is a neutral equilibrium: It remains at the allelic frequency it moved to when perturbed.

Stable, unstable, and neutral equilibrium points can be visualized as marbles in the bottom of a concave surface (stable), on the top of a convex surface (unstable), or on a level plane (neutral; fig. 20.1). Although more sophisticated mathematical formulas exist for determining whether an equilibrium is stable, unstable, or neutral, we will use graphical analysis for this purpose.

Figure 20.2 introduces the process of graphical analysis, which provides an understanding of the dynamics of an event or process by representing the event in graphical form. In figure 20.2, we have graphed equation 20.3, the $\Delta q$ equation of mutational dynamics. The ordinate, or y-axis, is $\Delta q$, the change in allelic frequency. The abscissa, or x-axis, is $q$, or allelic frequency. The diagonal line is the $\Delta q$ equation, the relationship between $\Delta q$ and $q$. Note that $\Delta q$ can be positive ($q$ is increasing) or negative ($q$ is decreasing), whereas $q$ is always positive (0–1.0). Graphical analysis can provide insights into the dynamics of many processes in population genetics.

The diagonal line in figure 20.2 crosses the $\Delta q = 0$ line at the equilibrium value ($\hat{q}$) of 0.167. This line also shows us the changes in allelic frequency that occur in a population not at the equilibrium point. We will look at two examples of populations under the influence of mutation pressure, but not at equilibrium: one at $q = 0.1$ (below equilibrium) and one at $q = 0.9$ (above equilibrium).

If we substitute $q = 0.1$ into equation 20.3, we get a $\Delta q$ value of $4 \times 10^{-6}$. If we substitute $q = 0.9$ into the equation, we get a $\Delta q$ value of $-4.4 \times 10^{-5}$. In other words, when the population is below equilibrium, $q$ increases ($\Delta q = +4 \times 10^{-6}$); if the population is above equilibrium, $q$ decreases ($\Delta q = -4.4 \times 10^{-5}$). We can read these same conclusions directly from the graph in figure 20.2.

We can see that the mutational equilibrium is a stable one. Any population whose allelic frequency is not at the equilibrium value tends to return to that equilibrium value. A shortcoming of this model is that it provides no obvious information revealing the time frame for reaching equilibrium. To derive the equations needed to determine this parameter is beyond our scope. (We could use computer simulation or integrate equation 20.3 with respect to time.) In a large population, any great change in allelic frequency caused by mutation pressure alone takes an extremely long time. Most mutation rates are on

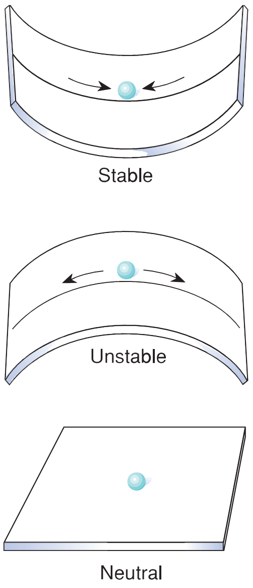


**Figure 20.1** Types of equilibria: stable, unstable, and neutral.

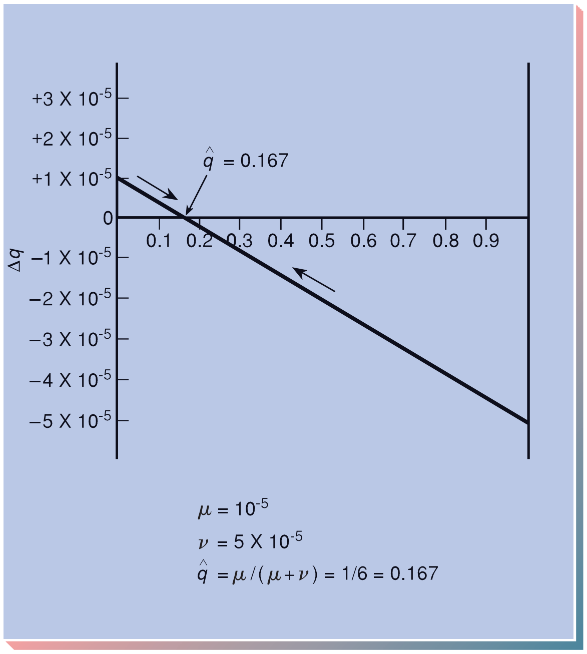


**Figure 20.2** Graphical analysis of mutational equilibrium. The graph of the mutational $\Delta q$ equation shows that when the population is perturbed from the equilibrium point ($q = 0.167$), it returns to that equilibrium point. At $q$ values above equilibrium, change is negative, tending to return the population to equilibrium. At $q$ values below equilibrium, change is positive, also tending to return the population to equilibrium.

the order of $10^{-5}$, and equation 20.3 shows that change will be very slow with values of this magnitude. For example, if $\mu = 10^{-5}$, $\nu = 10^{-6}$, and $p = q = 0.5$, $\Delta q = (0.5 \times 10^{-5}) - (0.5 \times 10^{-6}) = 4.5 \times 10^{-6}$, or 0.0000045. It usually takes thousands of generations to get near equilibrium, which is approached asymptotically.

As you can see from the low values of mutation rates, it would usually be nearly impossible to detect perturbations to the Hardy-Weinberg equilibrium by mutation in any one generation. The mutation rate can, however, determine the eventual allelic frequencies at equilibrium if no other factors act to perturb the gradual changes that mutation rates cause. Mutation can also affect final allelic frequencies when it restores alleles that natural selection is removing, a situation we will discuss at the end of the chapter. More important, mutation provides the alternative alleles that natural selection acts upon.

## MIGRATION

Migration is similar to mutation in the sense that it adds or removes alleles and thereby changes allelic frequencies. Human populations are frequently affected by migration.

Assume two populations, natives and migrants, both containing alleles *A* and *a* at the *A* locus, but at different frequencies ($p_N$ and $q_N$ versus $p_M$ and $q_M$), as shown in figure 20.3. Assume that a group of migrants joins the native population and that this group of migrants makes up a fraction *m* (e.g., 0.2) of the new conglomerate population. Thus, the old residents, or natives, will make up a proportionate fraction ($1 - m$; e.g., 0.8) of the combined population. The conglomerate *a*-allele frequency, $q_c$, will be the weighted average of the allelic frequencies of the natives and migrants (the allelic frequencies weighted—multiplied—by their proportions):

$$q_c = mq_M + (1 - m)q_N \qquad \textbf{(20.8)}$$

$$q_c = q_N + m(q_M - q_N) \qquad \textbf{(20.9)}$$

The change in allelic frequency, *a*, from before to after the migration event is

$$\Delta q = q_c - q_N = [q_N + m(q_M - q_N)] - q_N \qquad \textbf{(20.10)}$$

$$\Delta q = m(q_M - q_N) \qquad \textbf{(20.11)}$$

We then find the equilibrium value, $\hat{q}$ (at $\Delta q = 0$). Remembering that, in a product series, any multiplier with the value of zero makes the whole expression zero, $\Delta q$ will be zero when either

$$m = 0 \text{ or } q_M - q_N = 0; \; q_M = q_N$$

The conclusions we can draw from this model are intuitive. Migration can upset the Hardy-Weinberg equilibrium. Allelic frequencies in a population under the influence of migration will not change if either the size of the migrant group drops to zero (*m*, the proportion of the conglomerate made up of migrants, drops to zero) or the allelic frequencies in the migrant and resident groups become identical.

This migration model can be used to determine the degree to which alleles from one population have entered another population. It can analyze the allele interactions in any two populations. We can, for example, analyze the amount of admixture of alleles from Mongol populations with eastern European populations to explain the relatively high levels of blood type B in eastern European populations (if we make the relatively unrealistic assumption that each of these groups is homogeneous). The calculations are also based on a change happening all in one generation, which did not happen. Blood type and other loci can be used to determine allelic frequencies in western European, eastern European, and Mongol populations. We can rearrange equation 20.9 to solve for *m*, the proportion of migrants:

$$m = \frac{q_c - q_N}{q_M - q_N} \qquad \textbf{(20.12)}$$

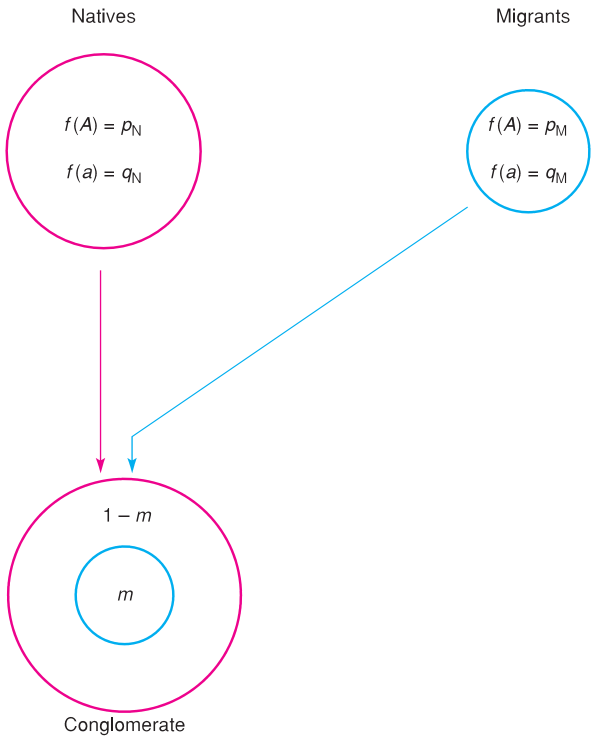


**Figure 20.3** Diagrammatic view of migration. A group of migrants enters a native population, making up a proportion, *m*, of the final conglomerate population.

From one sample, we find that the *B* allele is 0.10 in western Europe, taken as the resident or native population ($q_N$); 0.12 in eastern Europe, the conglomerate population ($q_C$); and 0.21 in Mongols, the migrants ($q_M$). Substituting these values into equation 20.12 gives a value for *m* of 0.18. That is, given the stated assumptions, 18% of the alleles in the eastern European population were brought in by genetic mixture with Mongols.

When a migrant group first joins a native group, before genetic mixing (mating) takes place, the Hardy-Weinberg equilibrium of the conglomerate population is perturbed, even though both subgroups are themselves in Hardy-Weinberg proportions. A decrease will occur in heterozygotes in the conglomerate population as compared to what we would predict from the allelic frequencies of that population (the average allelic frequencies of the two groups). This is a phenomenon of subdivision referred to as the **Wahlund effect.** The reason this happens is because the relative proportions of heterozygotes increase at intermediate allelic frequencies. As allelic frequencies rise above or fall below 0.5, the relative proportion of heterozygotes decreases.

In a conglomerate population, the allelic frequencies will be intermediate between the values of the two subgroups because of averaging. This generally means the predicted proportion of heterozygotes will be higher than the actual average proportion of heterozygotes in the two subgroups. An example is worked out in table 20.1. Assume that the two subgroups each make up 50% of the conglomerate population. In subgroup 1, $p = 0.1$ and $q = 0.9$; in subgroup 2, $p = 0.9$ and $q = 0.1$. Each subgroup will have 18% heterozygotes. The average, $(0.18 + 0.18)/2 = 0.18$, is the proportion of heterozygotes actually in the population. However, the conglomerate allelic frequencies are $p = 0.5$ and $q = 0.5$, leading to the expectation that 50% of the population will be heterozygotes. Hence, the observed frequency of heterozygotes is lower than the expected frequency (i.e., the Wahlund effect).

**Table 20.1** **The Wahlund Effect: Heterozygote Frequencies Are Below Expected in a Conglomerate Population**

| | Subgroup 1 | Subgroup 2 | Conglomerate | |
|---|---|---|---|---|
| $p$ | 0.1 | 0.9 | 0.5 | |
| $q$ | 0.9 | 0.1 | 0.5 | |
| | | | Expected | Observed |
| $p^2$ | 0.01 | 0.81 | 0.25 | 0.41 |
| $2pq$ | 0.18 | 0.18 | 0.50 | 0.18 |
| $q^2$ | 0.81 | 0.01 | 0.25 | 0.41 |

*Note:* In this example, the subgroups are of equal sizes.

We should note that the same logic holds even if both populations have allelic frequencies above or below 0.5. Also, this effect happens when an observer samples what he or she thinks is a single population but is actually a population subdivided into several demes. When most population geneticists sample a population and find a deficiency of heterozygotes, they first think of inbreeding and then of subdivision, the Wahlund effect. (A further complication is that inbreeding leads to subdivision, and subdivision leads to inbreeding. Statistics have been developed to try to separate the effects of these two phenomena.) As soon as random mating occurs in a subdivided population, Hardy-Weinberg equilibrium is established in one generation. We refer to a population in which the individuals are mating at random as unstructured or **panmictic.**

## SMALL POPULATION SIZE

Another variable that can upset the Hardy-Weinberg equilibrium is small population size. The Hardy-Weinberg equilibrium assumes an infinitely large population because, as defined, it is **deterministic,** not stochastic. That is, the Hardy-Weinberg equilibrium predicts exactly what the allelic and genotypic frequencies should be after one generation; it ignores variation due to sampling error. Obviously, every population of organisms on earth violates the Hardy-Weinberg assumption of infinite population size.

### *Sampling Error*

The zygotes of every generation are a sample of gametes from the parent generation. Sampling errors are the changes in allelic frequencies from one generation to the next that are due to inexact sampling of the alleles of the parent generation. Toss a coin one hundred times, and chances are, it will not land heads exactly fifty times. However, as the number of coin tosses increases, the percentage of heads will approach 50%, a percentage reached with certainty only after an infinite number of tosses. The same applies to any sampling problem, from drawing cards from a deck to drawing gametes from a gene pool.

If small population size is the only factor causing deviation from Hardy-Weinberg equilibrium, it will cause the allelic frequencies of a population to fluctuate from generation to generation in the process known as random genetic drift. In other words, an *Aa* heterozygote will sometimes produce several offspring that have only the *A* allele, or sometimes random mortality will kill a disproportionate number of *aa* homozygotes. In either case, the next generation may not have the same allelic frequencies as the

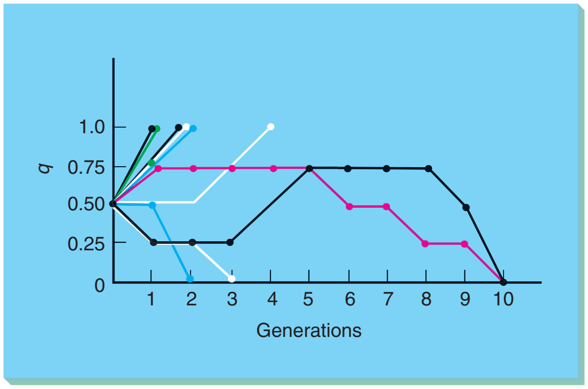


**Figure 20.4** Random genetic drift. Ten populations, each consisting of two individuals with initial $q = 0.5$, all go to fixation or loss of the *a* allele (four or zero copies) within ten generations due to the sampling error of gametes. Once the *a* allele has been fixed or lost, no further change in allelic frequency will occur (barring mutation or migration). We show a population of only two individuals to exaggerate the effects of random genetic drift.

present generation. The end result will be either fixation or loss of any given allele ($q = 1$ or $q = 0$; fig. 20.4), although which will be fixed or lost depends on the original allelic frequencies. The rate of approach to reach the fixation-loss endpoint depends on the size of the population.

## *Simulation of Random Genetic Drift*

We can investigate the process of random genetic drift mathematically by starting with a large number of populations of the same finite size and observing how the distribution of allelic frequencies among the populations changes in time due only to random genetic drift. For example, we can start with one thousand hypothetical populations, each containing one hundred individuals, with the frequency of the *a* allele, $q$, 0.5 in each (fig. 20.5). We measure time in generations, $t$, as a function of the population size, $N$ (one hundred in this example). For instance, $t = N$ is generation one hundred, $t = N/5$ is generation twenty, and $t = 3N$ is generation three hundred. Then, by using computer simulation (or the **Fokker-Planck equation,** which physicists use to describe diffusion processes such as Brownian motion), we generate the series of curves shown in figure 20.6. These curves show that as the number of generations increases, the populations begin to diverge from $q = 0.5$. Approximately the same number of populations go to $q$ values above 0.5 as go to $q$ values below 0.5. Therefore, the distribution spreads symmetrically. When the distribution of allelic frequencies reaches the sides of the graph, some populations become fixed for the *a* allele and some lose it. In a sense, the sides act as sinks:

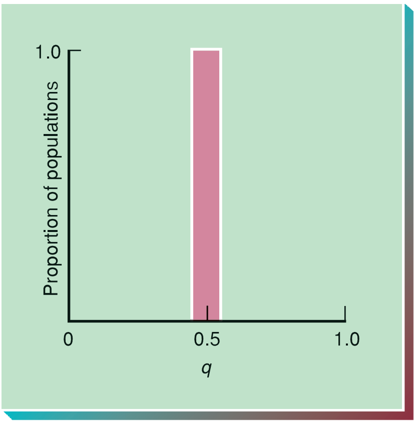


**Figure 20.5** Initial conditions of random drift model. One thousand populations, each of size one hundred, and each with an allelic frequency ($q$) of 0.5.

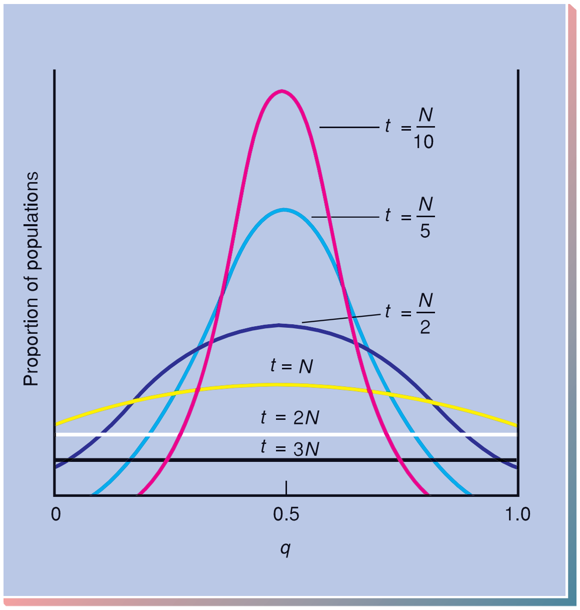


**Figure 20.6** Genetic drift in small populations: $q = 0.5$. After time passes, the populations of figure 20.5 begin to diverge in their allelic frequencies. Time is measured in population size ($N$), showing that the effects of random genetic drift are qualitatively similar in populations of all sizes; the only difference is the timescale. (From M. Kimura, "Solution of a process of random genetic drift with a continuous model," *Proceedings of the National Academy of Sciences, USA,* 41:144-50, 1955. Reprinted by permission.)

Any population that has the *a* allele lost or fixed will be permanently removed from the process of random genetic drift. Without mutation to bring one or the other allele

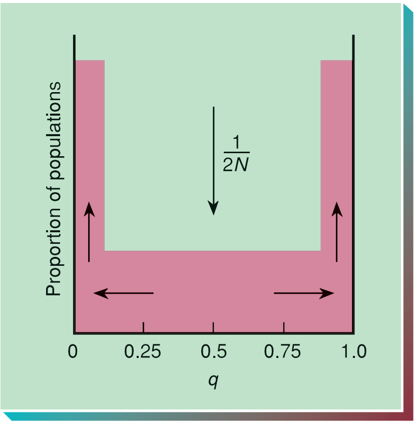


**Figure 20.7** Continued genetic drift in the one thousand populations, each numbering one hundred in size, shown in figures 20.5 and 20.6. After approximately 2*N* generations, the distribution is flat, and populations are going to loss or fixation of the *a* allele at a rate of 1/2*N* populations per generation. (From S. Wright, "Evolution in Mendelian Populations," *Genetics,* 97:114. Copyright © 1931 Genetics Society of America.)

back into the gene pool, these populations maintain a constant allelic frequency of zero or 1.0.

At a point between *N* (one hundred) and 2*N* (two hundred) generations, the distribution of allelic frequencies flattens out and begins to lose populations to the edges (fixation or loss) at a constant rate, as shown in figure 20.7. The rate of loss is about 1/2*N* (1/200), or 0.5% of the populations per generation. If the initial allelic frequency was not 0.5, everything is shifted in the distribution (fig. 20.8), but the basic process is the same—in all populations, sampling error causes allelic frequencies to drift toward fixation or elimination. If no other factor counteracts this drift, every population is destined to eventually be either fixed for or deficient in any given allele.

The amount of time the process takes depends on population size. The example used here was based on small populations of one hundred. If we substitute one million for one hundred in figure 20.6, a flat distribution of populations would not be reached for two million generations, rather than two hundred generations. Thus, a population experiences the effect of random genetic drift in inverse proportion to its size: Small populations rapidly fix or lose a given allele, whereas large populations take longer to show the same effects. Genetic drift also shows itself in several other ways.

## Founder Effects and Bottlenecks

Several well-known genetic phenomena are caused by populations starting at or proceeding through small numbers. When a population is initiated by a small, and therefore genetically unrepresentative, sample of the parent population, the genetic drift observed in the subpopulation is referred to as a **founder effect.** A classic human example is the population founded on Pitcairn Island by several of the *Bounty* mutineers and some Polynesians. The unique combination of Caucasian and Polynesian traits that characterizes today's Pitcairn Island population resulted from the small number of founders for the population.

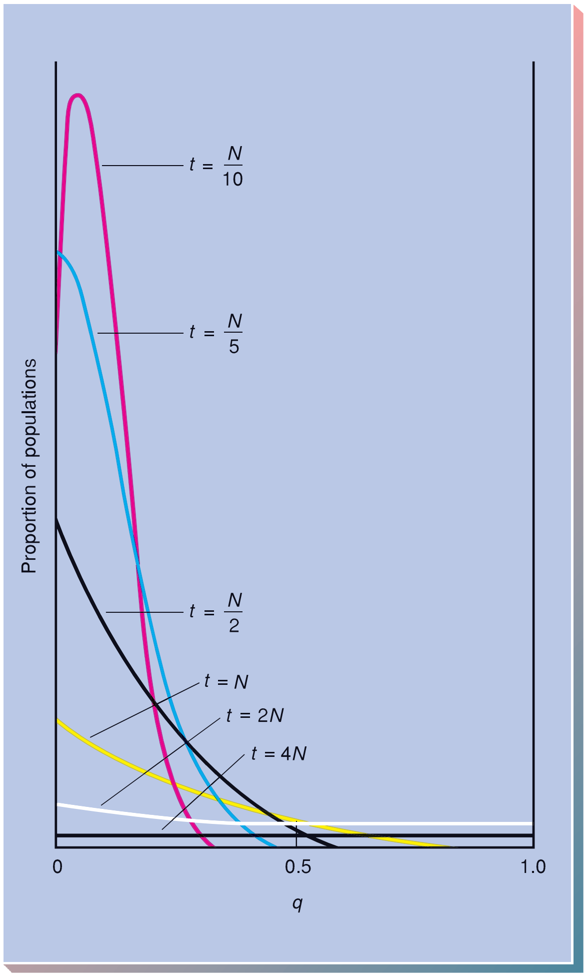


**Figure 20.8** Random genetic drift in small populations with $q = 0.1$. Compare this figure with figure 20.6. In this case, the probability of fixation of the *a* allele is 0.1, and the probability of its loss is 0.9. (From M. Kimura, "Solution of a process of random genetic drift with a continuous model," *Proceedings of the National Academy of Sciences, USA,* 41:144–50, 1955. Reprinted by permission.)

Sometimes populations go through **bottlenecks,** periods of very small population size, with predictable genetic results. After the bottleneck, the parents of the next generation have been reduced to a small number and may not be genetically representative of the original population. The field mice on Muskeget Island, Massachusetts, have a white forehead blaze of hair not commonly found in nearby mainland populations. Presumably, the island population went through a bottleneck at the turn of the century, when cats on the island reduced the number of mice to near zero. The population was reestablished by a small group of mice that happened by chance to contain several animals with this forehead blaze.

## NATURAL SELECTION

Although mutation, migration, and random genetic drift all influence allelic frequencies, they do not necessarily produce populations of individuals that are better adapted to their environments. Natural selection, however, tends to that end. The consequence of natural selection, Darwinian evolution, is considered in detail in the next chapter. We discuss here the algebra behind the process of natural selection. Artificial selection, as practiced by animal and plant breeders, follows the same rules.

### How Natural Selection Acts

**Selection,** or **natural selection,** is a process whereby one phenotype and, therefore, one genotype leaves relatively more offspring than another genotype, measured by both reproduction and survival. Selection is thus a matter of **reproductive success,** the relative contribution of that genotype to the next generation. It is important to remember that selection acts on whole organisms and thus on phenotypes. However, we analyze the process by looking directly at the genotype, usually only at one locus.

#### Fitness

A measure of reproductive success is the **fitness,** or **adaptive value,** of a genotype. A genotype that, compared with other genotypes, leaves relatively more offspring that survive to reproduce has the higher fitness. (Note that this use of the word *fitness* differs from our common notion of physical fitness.)

Fitness is usually computed to vary from zero to one (0–1) and is always related to a given population at a given time. For example, in a normal environment, fruit flies with long wings may be more fit than fruit flies with short wings. But in a very windy environment, a fruit fly with limited flying ability may survive better than one with the long-winged genotype, which will be blown around by the wind. Thus, fitness (usually assigned the letter $W$) is relative to a given circumstance. In a given environment, the genotype that leaves the most offspring is usually assigned a fitness of $W = 1$, and a lethal genotype has a fitness of $W = 0$. Any other genotype has a fitness value between zero and one. A number of factors can decrease this fitness value, $W$, below one. A **selection coefficient** measures the sum of forces acting to prevent reproductive success. It is usually represented by the letter $s$ or $t$ and is defined by the fitness equation

$$W = 1 - s \tag{20.13}$$

and

$$s = 1 - W \tag{20.14}$$

Thus, as the selection coefficient increases, fitness decreases, and vice versa.

#### Components of Fitness

Natural selection can act at any stage of the life cycle of an organism. It usually acts in one of four ways. (1) The reproductive success of a genotype can be affected by prenatal, juvenile, or adult survival. Differential survival of genotypes is referred to as viability selection or **zygotic selection.** (2) A heterozygote can produce gametes with differential success when one of its alleles fertilizes more often than the other. This is termed **gametic selection.** A well-studied case is the *t*-allele (tailless) locus in house mice; the gametes of as many as 95% of the heterozygous males of the *Tt* genotype carry the *t* allele. (This phenomenon is also referred to as **segregation distortion** or **meiotic drive.**) Selection can also take place in two areas of the reproductive segment of an organism's life cycle. (3) Some genotypes may mate more often than others (have greater mating success), resulting in **sexual selection.** Sexual selection usually occurs when members of the same sex compete for mates or when females have some form of choice. Adaptations for fighting, such as antlers in male elk, or displaying, such as the peacock's tail, are the results of sexual selection. (4) Finally, some genotypes may be more fertile than other genotypes, resulting in **fecundity selection.** The particular variable of the life cycle that selection acts upon is termed a **component of fitness.**

#### Effects of Selection

Figure 20.9 shows the three main ways that the sum total of selection can act. **Directional selection** works by continuously removing individuals from one end of the phenotypic (and therefore, presumably, genotypic) distribution (e.g., short-necked giraffes are removed). Removal means disappearance through death or failure to reproduce (genetic death). Thus, the mean is constantly shifted toward

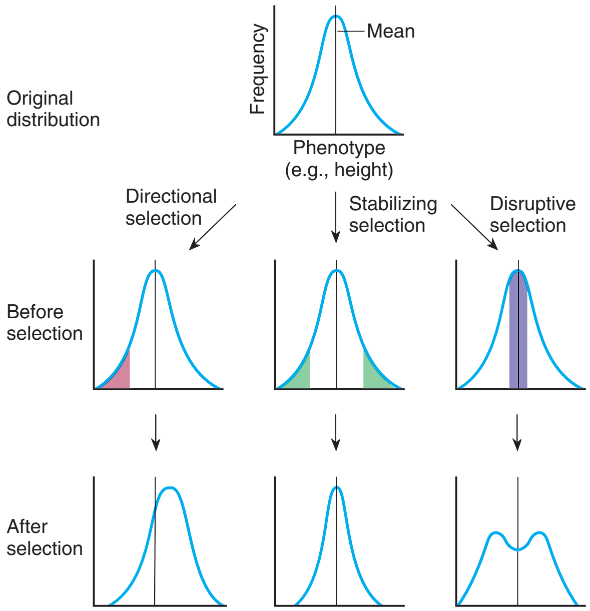


**Figure 20.9** Directional, stabilizing, and disruptive selection. *Colored areas* show the groups being selected against. At the *top* is the original distribution of individuals. The final distributions after selection appear in the *bottom row.*

the other end of the phenotypic distribution; in our example, the mean shifts toward long-necked giraffes. The evolution of neck length in giraffes, presumably by directional selection, has been documented from the geologic record.

**Stabilizing selection** (fig. 20.9) works by constantly removing individuals from both ends of a phenotypic distribution, thus maintaining the same mean over time. Stabilizing selection now works on giraffe neck length—it is neither increasing nor decreasing. **Disruptive selection** works by favoring individuals at both ends of a phenotypic distribution at the expense of individuals in the middle. It, like stabilizing selection, should maintain the same mean value for the phenotypic distribution. Disruptive selection has been carried out successfully in the laboratory for bristle number in *Drosophila.* Starting with a population with a mean number of sternopleural chaeta (bristles on one of the body plates) of about eighteen, investigators succeeded after twelve generations of getting a fly population with one peak of bristle numbers at about sixteen and another at about twenty-three (fig. 20.10).

## *Selection Against the Recessive Homozygote*

We can analyze selection by using our standard model-building protocol of population genetics—namely, define the initial conditions; allow selection to act; calculate the allelic frequency after selection ($q_{n+1}$); calculate $\Delta q$ (change in allelic frequency from one generation to the next); then calculate equilibrium frequency, $\hat{q}$, when $\Delta q$ becomes zero; and examine the stability of the equilibrium. In the analysis that follows, we consider a single autosomal locus in a diploid, sexually reproducing species with two alleles and assume that selection acts directly on the phenotypes in a simple fashion (i.e., it occurs at a single stage in the life of the organism, such as larval mortality in *Drosophila*). After selection, the individuals remaining within the population mate at random to form a new generation in Hardy-Weinberg proportions.

### *Selection Model*

In table 20.2, we outline the model for selection against the homozygous recessive genotype. The initial population is in Hardy-Weinberg equilibrium. Even with selection acting during the life cycle of the organism, Hardy-Weinberg proportions will be reestablished anew after each round of random mating, although presumably at new allelic frequencies. All selection models start out the same way. They diverge at the point of assigning fitness, which depends on the way natural selection is acting. In the model in table 20.2, the dominant homozygote and the heterozygote have

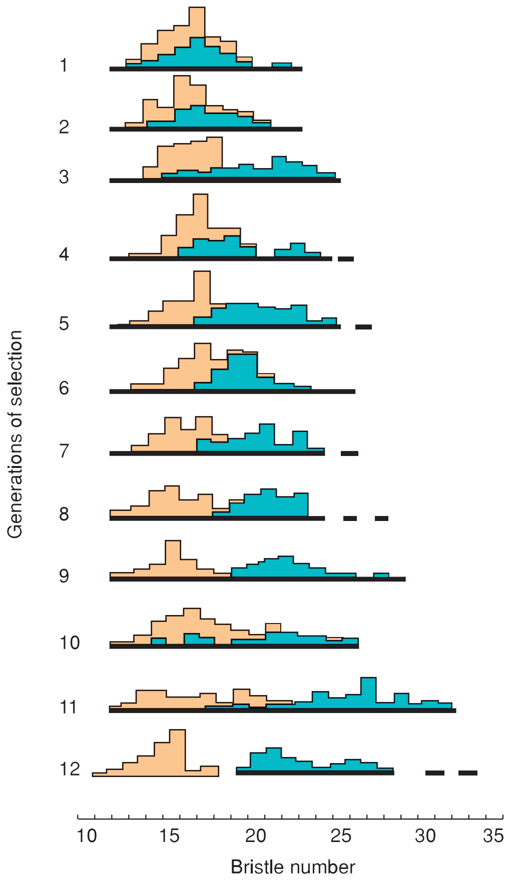


**Figure 20.10** Disruptive selection in *Drosophila melanogaster.* After twelve generations of selection for flies with either many or few bristles (chaetae) on the sternopleural plate, the population was bimodal. In other words, many flies in the population had either few or many bristles, but few flies had an intermediate bristle number. (Reprinted with permission from *Nature,* Vol. 193, J. M. Thoday and J. B. Gibson, "Isolation by Disruptive Selection." Copyright © 1962 Macmillan Magazines Limited.)

**Table 20.2** **Selection Against the Recessive Homozygote: One Locus with Two Alleles, *A* and *a***

| | Genotype | | | |
|---|---|---|---|---|
| | ***AA*** | ***Aa*** | ***aa*** | **Total** |
| Initial genotypic frequencies | $p^2$ | $2pq$ | $q^2$ | 1 |
| Fitness ($W$) | 1 | 1 | $1-s$ | |
| Ratio after selection | $p^2$ | $2pq$ | $q^2(1-s)$ | $1-sq^2 = \overline{W}$ |
| Genotypic frequencies after selection | $\frac{p^2}{\overline{W}}$ | $\frac{2pq}{\overline{W}}$ | $\frac{q^2(1-s)}{\overline{W}}$ | 1 |

the same fitness ($W = 1$). Natural selection cannot differentiate between the two genotypes because they both have the same phenotype. The recessive homozygote, however, is being selected against, which means that it has a lower fitness than the two other genotypes ($W = 1 - s$).

After selection, the ratio of the different genotypes is determined by multiplying their frequencies (Hardy-Weinberg proportions) by their fitnesses. The procedure follows from the definition of fitness, which in this case is a relative survival value. Thus, only $1 - s$ of the *aa* genotype survives for every one of the other two genotypes. For example, if $s$ were 0.4, then the fitness of the *aa* type would be $1 - s$, or 0.6. For every ten *AA* and *Aa* individuals that survive to reproduce, only six *aa* individuals would survive to reproduce. The total of the three genotypes after selection is $1 - sq^2$. That is,

$$p^2 + 2pq + q^2(1 - s) = p^2 + 2pq + q^2 - sq^2$$
$$= 1 - sq^2$$

### Mean Fitness of a Population

The value $(1 - sq^2)$ is referred to as the **mean fitness of the population, $\overline{W}$,** because it is the sum of the fitnesses of the genotypes multiplied (weighted) by the frequencies at which they occur. Thus, it is a weighted mean of the fitnesses, weighted by their frequencies. The new ratios of the three genotypes can be returned to genotypic frequencies by simply dividing by the mean fitness of the population, $\overline{W}$, as in the last line of table 20.2. (Remember that a set of numbers can be converted to proportions of unity by dividing them by their sum.) The new genotypic frequencies are thus the products of their original frequencies times their fitnesses, divided by the mean fitness of the population.

After selection, the new allelic frequency ($q_{n+1}$) is the proportion of *aa* homozygotes plus half the proportion of heterozygotes, or

$$q_{n+1} = \frac{q^2(1 - s)}{1 - sq^2} + \frac{pq}{1 - sq^2}$$
$$= \frac{q(q - sq + p)}{1 - sq^2}$$
$$= \frac{q(1 - sq)}{1 - sq^2} \qquad \textbf{(20.15)}$$

This model can be simplified somewhat if we assume that the *aa* genotype is lethal. Its fitness would be zero, and $s$, the selection coefficient, would be one. Equation 20.15 would then change to

$$q_{n+1} = \frac{q(1 - q)}{1 - q^2} \qquad \textbf{(20.16)}$$

Since $(1 - q^2)$ is factorable into $(1 - q)(1 + q)$, equation 20.16 becomes

$$q_{n+1} = \frac{q(1 - q)}{(1 - q)(1 + q)}$$

$$= \frac{q}{(1 + q)} \qquad \textbf{(20.17)}$$

The change in allelic frequency is then calculated as

$$\Delta q = q_{n+1} - q = \frac{q}{1 + q} - q$$

To solve this equation, $q$ is multiplied by $(1 + q)/(1 + q)$ so that both parts of the expression are over a common denominator:

$$\Delta q = \frac{q - q(1 + q)}{1 + q}$$

$$= \frac{-q^2}{1 + q} \qquad \textbf{(20.18)}$$

This is the expression for the change in allelic frequency caused by selection. Since selection will not act again until the same stage in the life cycle during the next generation, equation 20.18 is also an expression for the change in allelic frequency between generations.

Two facts should be apparent from equation 20.18. First, the frequency of the recessive allele ($q$) is declining, as indicated by the negative sign of the fraction. This fact should be intuitive because of the way selection was defined in the model (eliminating *aa* homozygotes). Second, the change in allelic frequency is proportional to $q^2$, which appears in the numerator of the expression. In other words, allelic frequency is declining as a relative function of the number of homozygous recessive individuals in the population. This fact is consistent with the premise of the selection model (with selection against the homozygous recessive genotype). This final formula supports the methodology of the model.

### Equilibrium Conditions

Next we calculate the equilibrium $q$ by setting the $\Delta q$ equation equal to zero, since a population in equilibrium will show no change in allelic frequencies from one generation to the next:

$$\frac{-q^2}{1 + q} = 0 \qquad \textbf{(20.19)}$$

For a fraction to be zero, the numerator must equal zero. Thus, $q^2 = 0$, and $\hat{q} = 0$. At equilibrium, the *a* allele should be entirely removed from the population. If the *aa* homozygotes are being removed, and if there is no mutation to return *a* alleles to the population, then eventually the *a* allele disappears from the population.

### Time Frame for Equilibrium

One shortcoming of this selection model is that it is not immediately apparent how many generations will be required to remove the *a* allele. The deficiency can be compensated for by using a computer simulation or by introducing a calculus differential into the model. Either method would produce the frequency-time graph of figure 20.11. This figure clearly shows that the *a* allele is removed more quickly when selection is stronger (when $s$ is larger) and that the curves appear to be asymptotic—the *a* allele is not immediately eliminated and would not be entirely removed until an infinitely large number of generations had passed. There is a reason for the asymptotic behavior of the graph: As the *a* allele becomes rarer and rarer, it tends to be found in heterozygotes (table 20.3). Since selection can remove only *aa* homozygotes, an *a* allele hidden in an *Aa* heterozygote will not be selected against. When $q = 0.5$, there are two heterozygotes for every *aa* homozygote. When

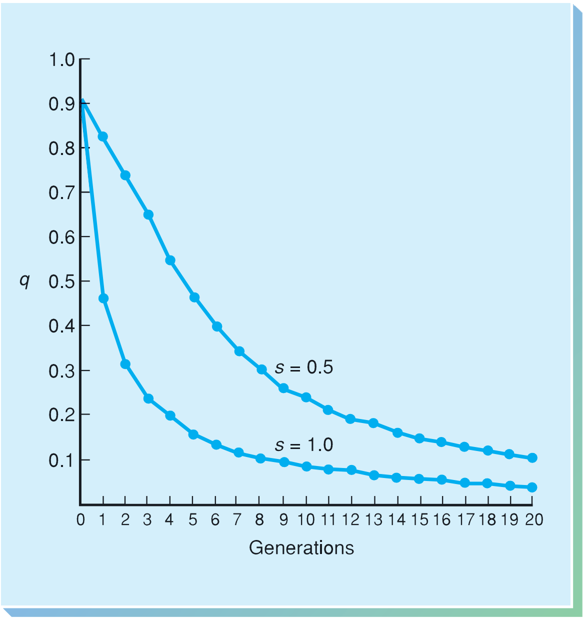


**Figure 20.11** Decline in $q$ (the frequency of the *a* allele) under different intensities of selection against the *aa* homozygote. Note that the loss of the *a* allele is asymptotic in both cases, but the drop in allelic frequency is more rapid with the larger selection coefficient.

**Table 20.3** **Relative Occurrence of Heterozygotes and Homozygotes as Allelic Frequency Declines: $q = f(a)$; $p = f(A)$**

| $q$ | $f(Aa)$ $(2pq)$ | $f(aa)$ $(q^2)$ | $f(Aa)/f(aa)$ |
|---|---|---|---|
| 0.5 | 0.50 | 0.25 | 2 |
| 0.2 | 0.32 | 0.04 | 8 |
| 0.1 | 0.18 | 0.01 | 18 |
| 0.01 | 0.0198 | 0.0001 | 198 |
| 0.001 | 0.001998 | 0.000001 | 1,998 |

$q = 0.001$, there are almost two thousand heterozygotes per *aa* homozygote. Remember, only the recessive homozygote is selected against. Natural selection cannot distinguish the dominant homozygote from the heterozygote.

## *Selection-Mutation Equilibrium*

Although a deleterious allele is eliminated slowly from a population, the time frame is so great that there is opportunity for mutation to bring the allele back. Given a population in which alleles are removed by selection and added by mutation, the point at which no change in allelic frequency occurs, the **selection-mutation equilibrium,** may be determined as follows. The new frequency ($q_{n+1}$) of the recessive *a* allele after nonlethal selection ($s < 1$) against the recessive homozygote is obtained by equation 20.15:

$$q_{n+1} = \frac{q(1 - sq)}{1 - sq^2}$$

Change in allelic frequency under this circumstance will thus be

$$\Delta q = q_{n+1} - q = \frac{q(1 - sq)}{(1 - sq^2)} - \frac{q(1 - sq^2)}{(1 - sq^2)}$$

$$= \frac{q - sq^2 - q + sq^3}{(1 - sq^2)}$$

$$= \frac{-sq^2(1 - q)}{1 - sq^2} \qquad \textbf{(20.20)}$$

Equation 20.20 is the general form of equation 20.18 for any value of *s*. The change in allelic frequency due to mutation can be found by using equation 20.4:

$$\Delta q = \mu p - \nu q$$

where $\mu$ and $\nu$ are the rate of forward and back mutation, respectively. When equilibrium exists, the change from selection will just balance the change from mutation. Thus,

$$\mu p - \nu q + \frac{-sq^2(1 - q)}{1 - sq^2} = 0$$

and

$$\mu p - \nu q = \frac{sq^2(1 - q)}{1 - sq^2} \qquad \textbf{(20.21)}$$

Now, some judicious simplifying is justified, because in a real situation, $q$ will be very small because the *a* allele is being selected against. Thus, $\nu q$ will be close to zero, and $1 - sq^2$ will be close to unity. Equation 20.21, therefore, becomes:

$$\mu p \cong sq^2(1 - q)$$

$$\mu(1 - q) \cong sq^2(1 - q)$$

$$q^2 \cong \mu/s$$

$$\hat{q} \cong \sqrt{\mu/s} \qquad \textbf{(20.22)}$$

In the case of a recessive lethal, *s* would be unity, so

$$q^2 \cong \mu \text{ and } \hat{q} \cong \sqrt{\mu}$$

If a recessive homozygote has a fitness of 0.5 ($s = 0.5$) and a mutation rate, $\mu$, of $1 \times 10^{-5}$, the allelic frequency at selection-mutation equilibrium will be

$$\hat{q} \cong \sqrt{\mu/s} \cong \sqrt{1 \times 10^{-5}/0.5} \cong \sqrt{2 \times 10^{-5}}$$

$$\cong 0.004$$

If the recessive phenotype were lethal, then

$$\hat{q} \cong \sqrt{\mu/s} \cong \sqrt{1 \times 10^{-5}/1}$$

$$\cong 0.003$$

These are very low equilibrium values for the *a* allele.

## *Types of Selection Models*

In view of the limited ways that fitnesses can be assigned, only a limited number of selection models are possible. Table 20.4 lists all possible selection models if we assume that fitnesses are constants and the highest fitness is one. (You might now go through the list of models and determine the equilibrium conditions for each.) Note that two possible fitness distributions are missing. There is no model in which fitnesses are $1 - s$, 1, and 1 for the $A_1A_1$, $A_1A_2$, and $A_2A_2$ genotypes, respectively (remembering that $p = f[A_1]$ and $q = f[A_2]$). That model is for selection against the $A_1A_1$ homozygote. Some reflection should show that this is the same model as model 1 of table 20.4, except that the $A_1$ allele is acting like a recessive allele. In other words, natural selection acts against $A_1A_1$ homozygotes, but not against the $A_1A_2$ and $A_2A_2$ genotypes. Thus, the model reduces to model 1 if we treat $A_1$ as the recessive allele and $A_2$ as the

**Table 20.4** **All Possible One-Locus, Two-Allele Selection Models (Assuming All Selection Coefficients Are Constants)**

| | Genotypic Fitness | | |
|---|---|---|---|
| Type of Selection | $A_1A_1$ | $A_1A_2$ | $A_2A_2$ |
| 1. Against recessive homozygotes | 1 | 1 | $1 - s$ |
| 2. Against heterozygotes | 1 | $1 - s$ | 1 |
| 3. Against one allele | 1 | $1 - s_1$ | $1 - s_2$ |
| 4. Against homozygotes | $1 - s_1$ | 1 | $1 - s_2$ |

dominant allele. Similarly, the $(1 - s_1, 1 - s_2, 1)$ model is eliminated for the same reason (allele $A_2$ is acting like the dominant allele and $A_1$ like the recessive allele). We now describe the outcome of each of the models in the table.

In both models 1 and 3 (table 20.4), selection is against genotypes containing the $A_2$ allele. Model 1, which we just derived in detail, is the model for a deleterious recessive allele. Almost any enzyme defect in a metabolic pathway fits this model, such as PKU, alkaptonuria, Tay-Sachs disease, and so on. In model 3, however, natural selection can detect the heterozygote, as is the case with deleterious alleles that are not completely recessive. An example would be the hemoglobin anomaly called thalassemia, a disorder common in some European and Asian populations, that produces a severe anemia in homozygotes and a milder anemia in heterozygotes. It should be clear that selection can more quickly eliminate a partially recessive allele than a completely recessive allele because the allele can no longer "hide" in the heterozygote.

Dominant or semidominant alleles (model 3) are usually more quickly removed from a population because they are completely open to selection. It takes an infinite number of generations to remove a recessive lethal allele, but only one generation for natural selection to remove a completely dominant lethal allele (see model 3, where $s_1 = s_2 = 1$). Examples of dominant deleterious traits in people are Huntington disease, facioscapular muscular dystrophy, and chondrodystrophy.

Model 2 is interesting because selection against the heterozygote leads to an unstable equilibrium at $q = 0.5$. If one heterozygote is removed by selection, one each of the two alleles is eliminated. However, if $p$ and $q$ are not equal (and thus not equal to 0.5), then one $A_1$ allele is not the same proportion of the $A_1$ alleles as one $A_2$ allele is of all the $A_2$ alleles. In other words, in a population of fifty individuals with $q = 0.1$ and $p = 0.9$, one $A_2$ allele is 10% (1/10) of the $A_2$ alleles, whereas one $A_1$ allele is only 1.1% (1/90) of the $A_1$ alleles. Removing one each of the two alleles causes a decrease in $q$. Therefore, a population following model 2 is at equilibrium at $p = q = 0.5$. However, this is an unstable equilibrium. Any perturbation that changes the allelic frequencies causes the rarer allele to be selected against and eventually removed from the population. An example is the maternal-fetal incompatibility at the Rh locus in human beings. The disease erythroblastosis occurs only in heterozygous fetuses ($Rh^+Rh^-$) in Rh-negative ($Rh^-Rh^-$) mothers. Heterozygotes are, therefore, selected against.

In model 4, selection is against homozygotes. This model is called the **heterozygote advantage,** and we will derive its equilibrium condition because the results are important to evolutionary theory (table 20.5). At equilibrium

$$\Delta q = \frac{pq(s_1p - s_2q)}{\overline{W}} \tag{20.23}$$

For this expression to be zero, either

$$p = 0, q = 0, \text{ or } (s_1p - s_2q) = 0$$

If $p = 0$ or $q = 0$, the result is trivial; the equilibrium exists only because of the absence of one of the alleles. The more meaningful equilibrium occurs when $s_1p - s_2q = 0$. In that case

$$s_1p = s_2q \text{ or } s_1(1 - q) = s_2q$$

and

$$\hat{q} = \frac{s_1}{s_1 + s_2} \tag{20.24}$$

**Table 20.5** **Selection Model of Heterozygote Advantage: The *A* Locus with $A_1$ and $A_2$ Alleles**

| | Genotype | | | |
|---|---|---|---|---|
| | $A_1A_1$ | $A_1A_2$ | $A_2A_2$ | Total |
| Initial genotypic frequencies | $p^2$ | $2pq$ | $q^2$ | 1 |
| Fitness ($W$) | $1 - s_1$ | 1 | $1 - s_2$ | |
| Ratio after selection | $p^2(1 - s_1)$ | $2pq$ | $q^2(1 - s_2)$ | $1 - s_1p^2 - s_2q^2 = \overline{W}$ |
| Genotypic frequencies after selection | $\frac{p^2(1 - s_1)}{\overline{W}}$ | $\frac{2pq}{\overline{W}}$ | $\frac{q^2(1 - s_2)}{\overline{W}}$ | 1 |

## BOX 20.1

## Experimental Methods

### *A General Computer Program to Simulate the Approach to Allelic Equilibrium Under Heterozygote Advantage*

It is surprising how much insight we can gain into the processes of population genetics by modeling them on a computer. The simple computer program presented here calculates changing allelic frequencies due to random mating when alleles at a locus are under a heterozygote-advantage selection regime. The program is written in the Microsoft® Visual Basic language. You can simulate any of the selection models described in chapter 20 by simply changing the variables. Also, this program can model many of the other processes discussed in this and the last chapter; usually, only a few lines need to be changed to look at an entirely different process. Other computer programs can substitute. Output should be graphed. The program should be rerun several times with various sets of values for the allelic frequencies and fitnesses. If the outcome isn't clear by twenty-five generations, the number of generations can be increased with a few small changes in the program.

In the computer program (fig. 1), $p$ is set to 0.9, $q$ is $1 - p$ (0.1), and the three fitnesses are named w11, w12, and w22 for the *AA*, *Aa*, and *aa* genotypes, respectively. In this case, w11 is set to 0.4, w12 to 1, and w22 to 0.6, a model of heterozygote advantage;

*continued*

```
Sub Command1_Click ()
    Static q(25)
    Static p(25)
    Picture1.Cls

'Set variables

    p(1) = .9
    w11 = .4
    w12 = 1
    w22 = .6
    q(1) = 1 - p(1)

'Calculate p and q values

    For i = 2 To 25
      wbar = p(i - 1) ^ 2 * w11 + 2 * p(i - 1) * q(i - 1) * w12 + q(i - 1) ^ 2 * w22
      q(i) = (q(i - 1) ^ 2 * w22 + p(i - 1) * q(i - 1) * w12) / wbar
      p(i) = 1 - q(i)
    Next i

'Draw axes and grid

    Picture1.Scale (-1, 1.1)-(26, -.1)
    Picture1.Line (0, 0)-(0, 1)
      For i = 0 To 10
         Picture1.Line (0, .1 * i)-(25, .1 * i)
      Next i
      For i = 5 To 25 Step 5
         Picture1.Line (i, 0)-(i, 1)
      Next i

'Draw q values

    Picture1.DrawWidth = 5
      For i = 1 To 25
         Picture1.PSet (i, q(i))
      Next i
    End Sub
```

**Figure 1** A Microsoft® Visual Basic computer program for the simulation of heterozygote advantage. The first statement indicates that the program is run by clicking a command button. Twenty-five values of $q$ and $p$ are calculated and stored for printing. The program also prints a grid of lines at increments of $q = 0.1$ and generations = 5.

## BOX 20.1 CONTINUED

the number of generations is twenty-five. The program calculates the mean fitness of the population, wbar, as $p^2$(w11) + $2pq$(w12) + $q^2$(w22); it then calculates the new allelic frequencies after one generation of selection (the new proportion of *aa* homozygotes plus half the proportion of heterozygotes). The program then repeats this process twenty-five times, storing each new *q* in the array *q*(i). The graphic output shown in figure 2 results. As you can see, *q* is approaching 0.6. If you would like to see the values generated by the program, appropriate print statements can be added.

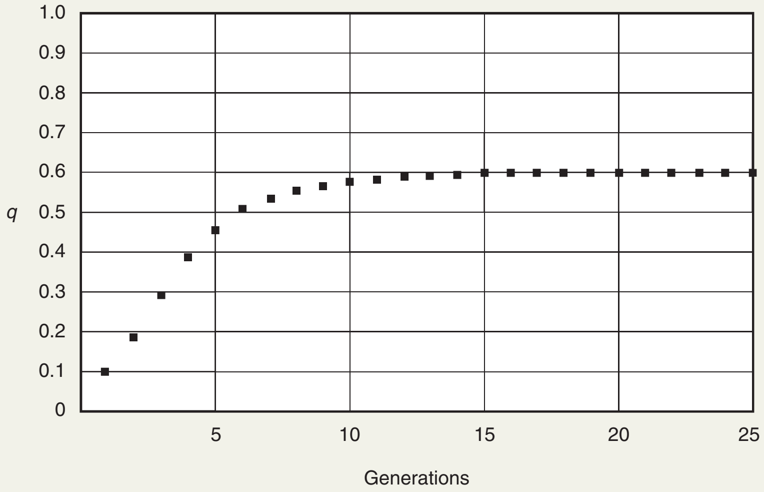


**Figure 2** Graphical output of the computer program from figure 1, with axis labels drawn in. The frequency of the *a* allele, *q*, begins at 0.1 and asymptotes toward 0.6.

Since $p + q = 1$,

$$\hat{p} = \frac{s_2}{s_1 + s_2} \qquad \textbf{(20.25)}$$

Several interesting conclusions follow. First, unlike the other models of selection, this model allows a population to maintain both alleles. We can demonstrate that this equilibrium is stable by graphing the $\Delta q$ value against $q$. Such a graph appears in figure 20.12, in which $q$ is the frequency of allele $A_2$ and the fitnesses of genotypes $A_1A_1$, $A_1A_2$, and $A_2A_2$ are assumed to be 0.8, 1, and 0.7, respectively. Note that if the equilibrium is perturbed by an increase or decrease in $q$, the population returns to the point of equilibrium. Second, the equilibrium is independent of the original allelic frequencies since it involves only the selection coefficients, $s_1$ and $s_2$. Last, the equilibrium for each allele (equations 20.24 and 20.25) is directly proportional to the selection coefficient against the other allele. As the selection against $A_1$ increases ($s_1$ increases), the equilibrium shifts toward a higher value of $q$ (more $A_2$ alleles; box 20.1).

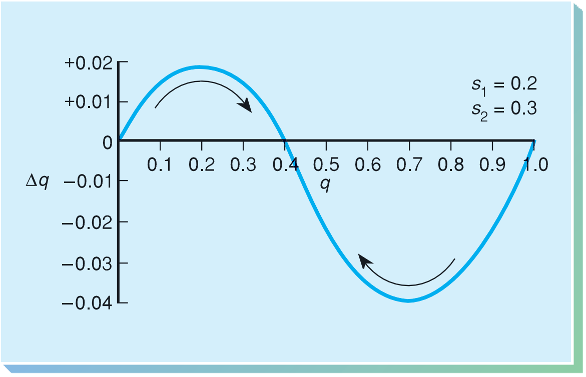


**Figure 20.12** Plot of allelic frequency ($q$) versus change in allelic frequency ($\Delta q$) for a polymorphism maintained by heterozygote advantage. In this case, $s_1 = 0.2$ and $s_2 = 0.3$; the equilibrium value, $\hat{q}$, is 0.4. When perturbed, the population tends to return to this value unless the perturbation brings $q$ to either 1.0 or 0.0, in which case the population is either fixed for the *a* allele or has lost it. In both cases, no further change in allelic frequency will take place, barring mutation or migration.

## SUMMARY

**STUDY OBJECTIVE 1:** To develop ways to analyze population genetics problems 571

A five-step protocol is presented to determine equilibrium allelic frequencies.

**STUDY OBJECTIVE 2:** To analyze the effects of mutation, migration, and population size on the Hardy-Weinberg equilibrium 571–577

The effects of relaxing some of the assumptions of the Hardy-Weinberg equilibrium are analyzed. Both mutation and migration transport alleles in and out of a population. Mutation provides the variability on which natural selection acts, but it usually does not directly affect the equilibrium because mutation rates are usually very low. If two randomly mating populations merge, or if two randomly mating demes are mistakenly treated as a single deme, the conglomerate will be deficient in heterozygotes. This deviation is called the Wahlund effect.

Finite population size is a source of sampling error. It results in changes in allelic frequencies known as random genetic drift. The smaller the population, the more rapidly allelic frequencies change. The dynamics of random genetic drift were studied graphically.

**STUDY OBJECTIVE 3:** To study the ways in which natural selection results in organisms adapted to their environments 577–584

Natural selection is defined by differential reproductive success. Depending upon which phenotypes are most fit, natural selection can act in several ways to change allelic and genotypic frequencies. Selection against the recessive homozygote acts to remove the allele from the population. Mutation brings the allele back into the population. Thus, a selection-mutation equilibrium maintains the unfavorable allele at a relatively low frequency. Heterozygote advantage maintains both alleles in a population.

## SOLVED PROBLEMS

**PROBLEM 1:** At a particular locus, there are two alleles, *B* and *b*. The mutation rate of *B* to *b* is $3.5 \times 10^{-4}$, whereas the mutation rate of *b* to *B* is $6 \times 10^{-8}$. What is the equilibrium frequency of the *b* allele, assuming no other factor is operating in this population to disturb the Hardy-Weinberg equilibrium?

***Answer:*** We let $q = f(b)$, $\mu = 3.5 \times 10^{-4}$, and $\nu = 6 \times 10^{-8}$. We then simply substitute $\mu$ and $\nu$ into equation 20.6:

$$\begin{aligned}\hat{q} &= \mu/(\mu + \nu) = 3.5 \times 10^{-4}/\\&\quad(3.5 \times 10^{-4} + 6 \times 10^{-8})\\&= 0.9998\end{aligned}$$

**PROBLEM 2:** Given a population of about one million cicadas with a frequency of the *a* allele at the *A* locus of 0.75, what is the probability that the *a* allele will be lost due to random genetic drift? How much longer will the possible loss of the allele take than the loss of the allele would take in a population of one thousand?

***Answer:*** Regardless of the size of a finite population, random genetic drift takes place. The probability of the loss of an allele with a frequency of 0.75 is 0.25; the probability of its fixation is 0.75 (see fig. 20.8). Since it is convenient to measure time (number of generations) within populations of finite size in units of population size, we can see that an event that takes *N* generations will be one thousand generations in the small population, but one million generations in the large population. Thus, random genetic drift occurs in the larger population at about one-thousandth the rate of the small population.

**PROBLEM 3:** In a laboratory colony of fruit flies, the fitnesses of the genotypes of an electrophoretic locus (malate dehydrogenase) are determined. Three genotypes, *FF*, *FS*, and *SS*, have fitnesses of 0.85, 1.0, and 0.6, respectively. What is the equilibrium frequency of the slow allele (*S*)?

***Answer:*** If the fitnesses of the three genotypes *FF, FS,* and *SS* are as given, then the locus is exhibiting heterozygote advantage with selection coefficients of the two homozygotes of $s_1 = 0.15\ (1 - 0.85)$ and $s_2 = 0.4\ (1 - 0.6)$. If $q$ is the frequency of the slow allele, then, using equation 20.24,

$$\hat{q} = s_1/(s_1 + s_2) = 0.15/(0.15 + 0.4) = 0.27$$

## EXERCISES AND PROBLEMS*

### MUTATION

**1.** Consider a locus with alleles *A* and *a* in a large, randomly mating population under the influence of mutation.

- **a.** If the mutation rate of *A* to *a* is $6 \times 10^{-5}$, and the back-mutation rate to *A* is $7 \times 10^{-7}$, what is the equilibrium frequency of *a?*
- **b.** If $q = 0.9$ in generation *n*, what would it be one generation later, under only the influence of mutation?

**2.** Derive an expression for mutation equilibrium when no back mutation is occurring.

**3.** Consider a population in which $p = 0.9$ and $q = 0.1$. If the forward mutation rate, $A \rightarrow a$, is $5 \times 10^{-5}$ and the reverse mutation rate, $a \rightarrow A$, is $2 \times 10^{-5}$, calculate the equilibrium frequency, $\hat{q}$, of the *a* allele.

**4.** If the forward mutation rate, $A \rightarrow a$, is five times the reverse mutation rate, what is the equilibrium frequency of the *a* allele?

### MIGRATION

**5.** The following data refer to the $R^{\circ}$ allele in the Rh blood system:

frequency in western Europeans = 0.62
frequency in eastern Europeans = 0.45
frequency in Mongols = 0.03

What is the total proportion of alleles that have entered the eastern European population?

**6.** Given the data from problem 1 of chapter 19, what factors could have caused the population to leave Hardy-Weinberg equilibrium? (*See also* SMALL POPULATION SIZE and NATURAL SELECTION)

**7.** In a population of nine hundred butterflies, the frequency ($p$) of the fast allele of the enzyme phosphoenol pyruvate is 0.6, and the frequency of the slow form ($q$) is 0.4. Ninety butterflies migrate to this population, and the migrants have a slow-allele frequency of 0.8. Calculate the allelic frequencies of the new population.

**8.** If the frequency of the *N* allele is 0.25 in a native population, 0.32 in a conglomerate population, and 0.4 in a migrant population, what percentage of the *N* alleles in the conglomerate population were derived from the migrant population?

**9.** In a particular population, the frequency of allele *t* was 0.25 in a migrant population and 0.45 in the conglomerate population. If the migration rate was 0.1, calculate the frequency of *t* in the original, native population.

### SMALL POPULATION SIZE

**10.** In a population of five hundred individuals with a frequency of allele *A* of 0.7, what is the ultimate fate of the *A* allele? What is the probability that the population will eventually lose the *A* allele? How many are $N/5$ generations? $4N$ generations?

### NATURAL SELECTION

**11.** Differentiate among stabilizing, directional, and disruptive selection.

**12.** Derive a model of selection in which the fitness of the heterozygote is half the fitness of one of the homozygotes and twice the fitness of the other. Give expressions for the following:

- **a.** Mean population fitness
- **b.** Equilibrium allelic frequency (stable?)

**13.** Derive an expression for the equilibrium allelic frequencies under a model in which selection acts against heterozygotes. Is the equilibrium stable?

**14.** Table 20.6 describes selection at the *A* locus in a given diploid species in which $p = f(A)$ and $q = f(a)$.

- **a.** Describe the type of selection occurring here. Why does the total equal one before selection but $\overline{W}$, after?
- **b.** Derive an equation for *q* after one generation of selection ($q_{n+1}$).
- **c.** This system will reach equilibrium, with $\hat{p} = s_2/(s_1 + s_2)$. If selection is twice as strong against *aa* as against *AA*, what are the equilibrium allelic frequencies? If $s_1 = 0.1$ and $s_2 = 0.3$, what percentage of heterozygotes is at equilibrium?

**15.** Given a locus with alleles *A* and *a* in a sexually reproducing, diploid population in Hardy-Weinberg equilibrium, set up a model and the initial formula for the frequency of the dominant allele after one generation ($p_n$+1) if selection acts against the dominant phenotype. What are the equilibrium conditions?

**16.** There is a locus with alleles *A* and *a* in a large, randomly mating, diploid, sexually reproducing population. Allele *A* mutates to *a* at a rate of $\mu$, and no back mutation takes place. However, the *aa* homozygote is selected against with a fitness of $1 - s$. Give a formula for the equilibrium condition. If $\mu = 5 \times 10^{-5}$

---

* Answers to selected Exercises and Problems are on page A-22.

**Table 20.6**

| | Genotypes | | | |
|---|---|---|---|---|
| | $AA$ | $Aa$ | $aa$ | Total |
| Before selection | $p^2$ | $2pq$ | $q^2$ | 1 |
| Fitness ($W$) | $1 - s_1$ | 1 | $1 - s_2$ | |
| After selection | $p^2(1 - s_1)$ | $2pq$ | $q^2(1 - s_2)$ | $\overline{W} = 1 - s_1p^2 - s_2q^2$ |

and $s = 0.15$, what are the equilibrium allelic frequencies?

17. If a locus has alleles $A_1$ and $A_2$, what is the equilibrium frequency of $A_1$ if both homozygotes are lethal?

18. The following data were collected from a population of *Drosophila* segregating sepia ($s$) and wild-type ($s^+$) eye colors. One sample was taken when the eggs were deposited, and another was taken later among adults. Reconstruct the mode of selection.

| | $s^+s^+$ | $s^+s$ | $ss$ |
|---|---|---|---|
| Egg | 25 | 50 | 25 |
| Adult | 30 | 60 | 10 |

19. The data in table 20.7 come from T. Dobzhansky's work with chromosomal inversions in *Drosophila pseudoobscura*. They represent four samples from various altitudes in the Sierra Nevada Mountains in California. What would you say about, and what would you do in the lab to determine, the fitnesses of the inversions? What factors could cause the changes in fitness?

20. In a particular population with two alleles at a locus, the frequency of $AA$ individuals $= 0.25$, $Aa = 0.5$, and $aa = 0.25$. If the $AA$ genotype fitness $= 1$, $Aa = 0.8$, and $aa = 0.6$, what will the frequencies of $A$ and $a$ be in the next generation? Assume mutations do not occur.

**Table 20.7** **Data from Dobzhansky's Work**

| Elevation of Sample | Inversion | | | Others |
|---|---|---|---|---|
| | *ST* | *AR* | *CH* | |
| 6,800 ft | 26 | 44 | 16 | 14 |
| 4,600 ft | 32 | 37 | 19 | 12 |
| 3,000 ft | 41 | 35 | 14 | 10 |
| 800 ft | 46 | 25 | 16 | 13 |

*Note*: *ST* = standard; *AR* = Arrowhead; *CH* = Chiricahua

21. Calculate the frequency of the recessive $b$ allele in a population one generation after selection if in the original population $q = f(b) = 0.7$ and the relative fitness of $bb$ homozygotes is 0.4.

22. A type of dwarfism in dogs is caused by a recessive allele. The mutation rate from the normal to the mutant allele has been estimated at $5 \times 10^{-5}$, and the fitness of the dwarf is 0.2 when compared with normal individuals. Calculate the equilibrium frequency of the dwarf allele.

23. A recessive allele ($q = 0.5$) was initially neutral, but suddenly the environment changed and the recessive homozygote became lethal. What is $q$ one generation after selection begins? What is the expected frequency of the recessive allele two generations after selection?

## CRITICAL THINKING QUESTIONS

1. If the selection model of heterozygous disadvantage leads to the elimination of the rarer allele, why would such systems (e.g., the Rh blood system) still exist?

2. A scientist studied the distribution of electrophoretic genotypes in a sample of an insect species and found a deficiency of heterozygotes. How could this come about?

*Suggested Readings for chapter 20 are on page B-20.*

# 21

# EVOLUTION AND SPECIATION

## STUDY OBJECTIVES

1. To analyze the mechanisms of evolution and speciation 589
2. To investigate the mechanisms of the maintenance of genetic variation in natural populations, both selective and neutral 596
3. To discuss sociobiology, the evolution of social behavior 603

## STUDY OUTLINE

**Darwinian Evolution** 589
**Evolution and Speciation** 589
  Mechanisms of Cladogenesis 592
  Phyletic Gradualism Versus Punctuated Equilibrium 594
**Genetic Variation** 596
  Maintaining Polymorphisms 596
  Maintaining Many Polymorphisms 598
  Which Hypothesis Is Correct? 599
  Grand Patterns of Variation 600
**Sociobiology** 603
  Altruism 603
  Kin Selection and Inclusive Fitness 605
*Summary* 607
*Solved Problems* 607
*Exercises and Problems* 608
*Critical Thinking Questions* 609
**Box 21.1 Attacks on Darwinism** 590
**Box 21.2 Mimicry** 604

The cactus ground-finch (*Geospiza scandens*) from Santa Cruz Island, Galápagos. (© Frans Lanting/ Photo Researchers, Inc.)

Populations change, or evolve, through natural selection and the other forces that perturb the Hardy-Weinberg equilibrium. The merger of population genetics theory with the classical Darwinian view of evolution is known as **neo-Darwinism,** or the "new synthesis." In the two previous chapters, we laid the theoretical groundwork for an understanding of the process of evolution in natural populations. In this chapter, we concern ourselves with long-term evolution and speciation.

## DARWINIAN EVOLUTION

Charles Darwin (fig. 21.1) was a British naturalist who published his theory of evolution in 1859 in a book entitled *The Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life.* This book provided overwhelming support for evolution as well as a mechanism for the process. Darwin had been greatly influenced by the writings of the Reverend Thomas Malthus, who is best known for his theory that populations increase exponentially, whereas their food supplies increase arithmetically. Malthus, who proposed his theory in *An Essay on the Principle of Population* in 1798, was referring specifically to human populations and was trying to encourage people to reduce their birthrate rather than let their offspring starve to death. Malthus's writings impressed upon Darwin the realization that under limited resources—the usual circumstance in nature—not all organisms survive. In nature, organisms compete for the resources needed to survive.

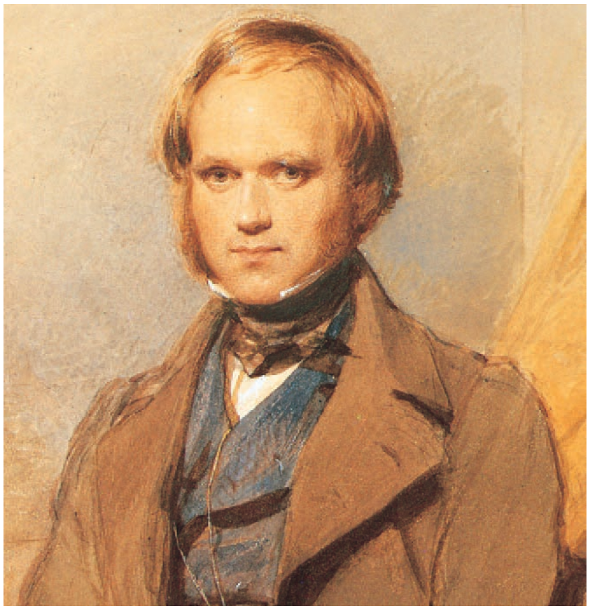


**Figure 21.1** Charles Darwin (1809–1882). Darwin was an English naturalist who first established the theory of organic evolution by natural selection. (Painting by George Richmond, 1840. Downe House, Downe, Kent. © Archiv/Photo Researchers, Inc.)

Darwin sailed aboard the HMS *Beagle,* a ship that circled the world from 1831 to 1836 with the primary purpose of charting the coast of South America. During his travels on the *Beagle,* Darwin amassed great quantities of observations (especially on South America and the Galápagos Islands) that led him to suggest a theory. Darwin proposed that organisms become adapted to their environment by the process of natural selection. In outline, the process works according to the following principles:

1. *Variation is a characteristic of virtually every group of animals and plants.* Darwin saw variation as an inherent property among individuals of all populations.
2. *Every group of organisms overproduces offspring.* Most populations maintain a relatively constant density over time. Thus, every parent, on average, just replaces itself. Therefore, most of the offspring the individuals of a population produce will die before they reproduce. Hence, in every group of organisms, there is an overabundance of young.
3. *Those that do survive and reproduce will pass on their genes in greater proportion.* This step is the cornerstone and the best-known part of Darwin's theory. Among all the organisms competing for a limited array of resources, only the organisms best able to obtain and utilize these resources survive **(survival of the fittest).** If the favorable characteristics of these individuals are inherited, these traits pass on to the next generation. These organisms then have the greatest reproductive success (box 21.1).

Thus, over time, if advantageous mutations arise, or if the environment changes, the characteristics of a population should change through the process of natural selection (directional or disruptive selection). A particularly well-adapted population in a stable environment may maintain its numbers through the forces of stabilizing selection (see fig. 20.9). Nonrandom mating, genetic drift, and migration may also play a role in population differentiation.

## EVOLUTION AND SPECIATION

The term **evolution** describes a change in genotypic frequencies, which usually results in a population of individuals better adapted to the environment than their ancestors were. **Speciation** comes in two different forms. (1) It may be the evolution of a population over time until the current population cannot be classified as

### BOX 21.1

## Ethics and Genetics

### *Attacks on Darwinism*

From time to time, attacks on neo-Darwinism are mounted, usually by persons who either view evolutionary theory as antireligious or who misunderstand Darwin's theory. One attack, entitled "Darwin's Mistake," by Tom Bethell, was published in *Harper's* magazine in 1976.

Bethell began by pointing out that Darwinian theory is a tautology rather than a predictive theory. (The term *tautology* means a statement that is true by definition.) That is, evolution is the survival of the fittest. But who are the fittest? Obviously, the individuals who survive. Thus, without an independent criterion for fitness other than survival, we are left with the statement that evolution is the survival of the survivors. This, indeed, is a tautology. But it is possible to assign independent criteria for fitness. Darwin wrote extensively about artificial selection in pigeons, in which the breeders' choice was the criterion for fitness. (Many novel breeds of pigeon have been created this way.) Plant and animal breeders have practiced artificial selection extensively. Here, survival is not the criterion for fitness; productivity is.

It is more difficult to establish *a priori* independent criteria of fitness in nature. Often, uncontrolled or unseen vagaries have major impacts on the course of events. Surely the temperature became colder before the mammoths became woolly. Is it then reasonable to predict that elephants would get woolly if the climate became colder in Africa today? The answer is no, for several reasons. First, the elephants might adapt to colder weather in any of a large number of different ways—they could get fatter, they could migrate, and so on. To some extent, adaptation depends not only on the changing environment, but also on the reserve variation within the gene pool of the species. Second, the elephants could become extinct; they might not be able to adapt at all. And third, if the climatic changes were not severe, the elephants might not change at all.

Predicting the exact course of evolution is nearly impossible. To provide independent criteria for fitness in nature is, therefore, very difficult. Some modern evolutionary biologists, although not doubting neo-Darwinism, do worry to some extent about the difficulties in testing modern evolutionary theory. However, lower-level experiments are done to test various aspects of evolution in specific systems. For example, in 1993, B. Grant and P. Grant hypothesized that changes in bill size would occur in the finch *Geospiza fortis* (see fig. 21.6) because of changing food size due to changing weather on the Galápagos Islands. Their proposal seems to be correct. In addition, the support for Darwinism (the fossil

belonging to the same **species** as the original population. This process is known as **anagenesis,** or **phyletic evolution** (*an* is Latin for without, *genesis* is Latin for birth or creation). (2) Speciation may also be the divergence of a population into two distinct forms (species) that exist simultaneously. This branching process is known as **cladogenesis** (*clado* is Greek for branch; fig. 21.2). What do we mean by the term *species?*

Before Darwin's time, **typological thinking** prevailed, and a species was defined as a group of organisms that were morphologically similar. All variants were considered imperfections of the model or type. One of Darwin's greatest contributions to modern biological theory was to treat variation as a normal phenomenon in a group of organisms. The modern **biological species concept** groups together as members of the same species organisms that can potentially interbreed. A species, therefore, is a group of organisms that can mate among themselves to produce fertile offspring.

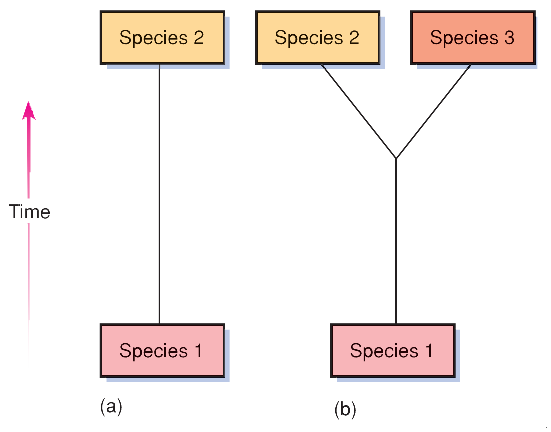


**Figure 21.2** Forms of speciation. In anagenesis (*a*), a species changes over time until it is so different from its progenitor that it is classified as a new species. In cladogenesis (*b*), speciation takes place as a branching process wherein one species becomes two or more.

record, embryology, comparative anatomy, geographic distributions, etc.) is so overwhelming that the general nature of evolution is not in doubt. We can clearly trace its path, although we cannot make exact predictions for its future.

From a philosophical point of view, neo-Darwinism is the general paradigm (broad concept) defining "normal" biology. Every scientific endeavor works under the umbrella of a paradigm. When enough inconsistencies appear, a new paradigm is sought to replace the old in what Thomas Kuhn called a "scientific revolution." In physics, relativity overthrew Newtonian principles. In biology, Darwinism overthrew the concept of a recent, biblically described origin of animals and plants. Darwinism became the paradigm because it explained many things in a consistent fashion that a recent origin of all forms of life could not. Neo-Darwinism will remain the current paradigm unless it is overthrown by a better theory that explains previous inconsistencies. To date, no major inconsistencies suggest that neo-Darwinism is not correct.

In his article, Bethell went on to try to refute neo-Darwinism using the following argument: Survival of the fittest can be redefined to mean that some organisms have more offspring than others. Thus, natural selection cannot be a creative force because the only thing it works on is organisms alive now, some having more offspring than others. How, asks Bethell, can this possibly give us tigers and horses from ancestors that did not look like tigers and horses? The answer is that mutation produces variants in the population. The organism best able to compete will leave the most offspring. With an array of different genotypes in a population, natural selection determines which genotypes will increase in future generations. Traits that give the bearer an advantage increase in the population, and evolution takes place. Natural selection was the force behind the evolution from the small Eocene horse to the modern *Equus.*

Misinterpretation of mutation is the basis for other attacks on Darwinism. For example, Darwinian evolution has been attacked as not feasible, since most mutations are deleterious. How, the argument goes, can evolution proceed by a combination of deleterious events? The answer is that although most mutations are deleterious, some are not. This is especially true in changing environments; yesterday's deleterious mutant may be today's favored mutant.

The most recent attacks on Darwinism have been launched by creationists, who have attempted to pass laws in many states requiring schools to teach the biblical version of creation as an alternative to Darwinism. The courts have rejected this position because creationism is not a scientific theory. It does not follow the rules of the scientific method wherein empirical evidence can refute it.

Unfortunately, the definition of species on the basis of interbreeding cannot be used in many places, mostly due to the technical problems of applying it. Taxonomists and paleontologists, who often use nonliving specimens (preserved or fossilized), use the **morphological species concept** as a working definition. Under this concept, two organisms are classified as belonging to the same species if they are morphologically similar. They are classified as belonging to two different species if they are as different as two organisms belonging to two recognized species. Other problems arise for taxonomists since speciation is a dynamic process. For example, isolated subgroups of a population may be in various stages of becoming new species; the rate of successful interbreeding among individuals from these subgroups may range from 0 to 100%. How should the in-betweens be classified? There is no correct answer. It depends on the circumstances.

Still other problems make it necessary to turn to the morphological species concept. Haploid and asexual species are hard to classify. Also, two organisms that will not interbreed in nature may do so in a laboratory setting. Thus, the interbreeding test carried out in the laboratory (as is done frequently) is not necessarily an adequate criterion for speciation. Other problems arise in classifying groups that are geographically isolated from each other, such as populations on islands. These individuals are physically isolated, but in many cases they can interbreed freely when brought together with their mainland counterparts. So, although there is a good theoretical definition of a species (potentially interbreeding individuals), more often than not it is necessary for biologists to apply the morphological species concept to determine whether two populations belong to the same species. In some cases, no decision can be made about the species status of a population. It is clear that a population has evolved, but it is not clear whether it has evolved enough to be called a new species. However, this is more of a problem for taxonomists and evolutionary biologists than for the organisms themselves.

## *Mechanisms of Cladogenesis*

### *Reproductive Isolation*

How does one species become two? Basically, **reproductive isolating mechanisms** must evolve to prevent two subpopulations from interbreeding when they come into contact. Reproductive isolating mechanisms are environmental, behavioral, mechanical, and physiological barriers that prevent individuals of two species from producing viable offspring. Following is a modification of the classification system of isolating mechanisms suggested by evolutionary biologist G. L. Stebbins:

1. *Prezygotic mechanisms* prevent fertilization and zygote formation.
   a. Residential—The populations live in the same region, but occupy different habitats.
   b. Seasonal or temporal—The populations exist in the same region, but are sexually mature at different times.
   c. Ethological (in animals only)—The populations are isolated by incompatible premating behavior.
   d. Mechanical—Cross-fertilization is prevented or restricted by incompatible differences in reproductive structures.
2. *Postzygotic mechanisms* affect the hybrid zygotes after fertilization has taken place.
   a. $F_1$ hybrid breakdown—$F_1$ hybrids are inviable or weak.
   b. Developmental hybrid sterility—Hybrids are sterile because gonads develop abnormally or because meiosis breaks down before it is completed.
   c. Segregational hybrid sterility—Hybrids are sterile because of abnormal distribution to the gametes of whole chromosomes, chromosome segments, or combinations of genes.
   d. $F_2$ breakdown—$F_1$ hybrids are normal, vigorous, and fertile, but the $F_2$ generation contains many weak or sterile individuals.

### *Allopatric, Parapatric, and Sympatric Speciation*

Reproductive isolating mechanisms are barriers to **gene flow,** the spread of genes between populations. These isolating mechanisms can evolve in three different ways, each of which defines a different mechanism of speciation. Usually, the mode of speciation is dictated by both the properties of the genetic systems of the organisms and stochastic (random) or accidental events. For example, vertebrates tend to have different speciation modes than phytophagous (plant-feeding) insects.

The appearance of a geographic barrier, such as a river or mountain, through the range of a species physically isolates populations of the species. Physical isolation can also occur if migrants cross a particular barrier and begin a new population (founder effect). The physically isolated populations can then evolve independently. If reproductive isolating mechanisms evolve, then two distinct species are formed, and if they come together in the future, they remain distinct species. Speciation that occurs because reproductive isolating mechanisms evolve during physical separation of the populations is called **allopatric speciation** (fig. 21.3). As evolutionary biologist Guy Bush pointed out, "Although examples in nature are difficult to substantiate . . . it [allopatric speciation] has been convincingly demonstrated in frogs . . . and lizards."

Reproductive isolating mechanisms usually originate incidentally to the speciation process. That is, they arise incidentally during the process of evolution in isolated populations rather than being selected for. When isolated populations come together again, incomplete isolating mechanisms may allow hybrids to form. If the hybrids are normal and viable and can freely interbreed with individuals of each parent population, then no speciation has taken place. However, if the hybrids are at a disadvantage, natural selection may favor stronger isolating mechanisms. In this case, organisms that mate with individuals from the other population leave fewer offspring. The result is a more effective barrier to hybridization. Regions in which previously isolated populations come into contact and produce hybrids are called **hybrid zones.**

Until recently, evolutionary biologists believed that allopatric speciation was the general rule. Many now believe that two other modes of speciation may occur frequently in certain groups of organisms. **Parapatric speciation** occurs when a population of a species that occupies a large range enters a new niche or habitat (fig. 21.3). Although no physical barrier arises, the new niche acts as a barrier to gene flow between the population in the new niche and the rest of the species. Here again, reproductive isolating mechanisms evolve to produce two species where there was only one before. Parapatric speciation is believed to have occurred often in relatively nonvagile animals such as snails, flightless grasshoppers, and annual plants. **Sympatric speciation** occurs when a polymorphism, which is the occurrence of alternative phenotypes in the same population, arises within an interbreeding population before a shift to a new niche. This mode of speciation may be common in parasites and phytophagous insects. For example, if a polymorphism arises within a parasitic species that allows an individual with a certain genotype to adapt to a new host, this genotype may be the forerunner of a new species. If the parasite not only feeds on the new host but also mates on the new host, a barrier to gene flow arises, although the parasite may be surrounded by other members of its species with the original genotype. Sympatric speciation can thus occur in the middle of a species range rather than at the edges (fig. 21.3).

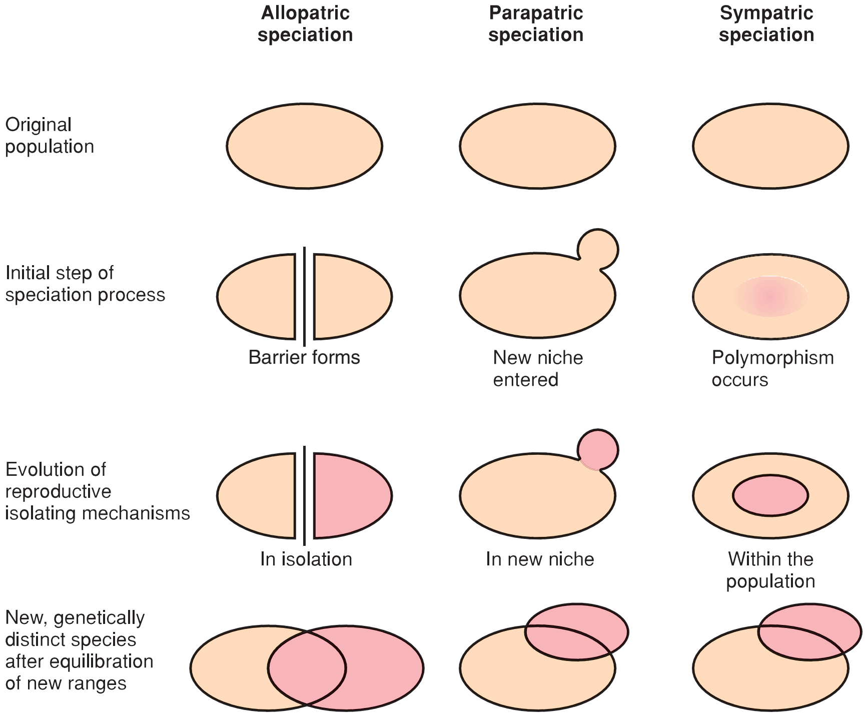

Allopatric speciation
Parapatric speciation
Sympatric speciation
Original population
Initial step of speciation process
Barrier forms
New niche entered
Polymorphism occurs
Evolution of reproductive isolating mechanisms
In isolation
In new niche
Within the population
New, genetically distinct species after equilibration of new ranges

**Figure 21.3** The three general mechanisms of speciation. In allopatric speciation, reproductive isolation evolves after the population has been geographically divided. In parapatric speciation, reproductive isolation evolves when a segment of the population enters a new niche. In sympatric speciation, reproductive isolation evolves while the incipient group is still in the vicinity of the parent population.

An example of incipient sympatric speciation has been seen recently in host races of the apple maggot fly (*Rhagoletis pomonella*) in North America (fig. 21.4). This fly was found originally only on hawthorn plants. However, in the nineteenth century, it spread as a pest to newly introduced apple trees. In fact, races are now known on pear and cherry trees and on rose bushes. These races have developed genetic, behavioral, and ecological differences from the original hawthorn-dwelling parent. Evolutionary biologists view this as an opportunity to observe sympatric speciation as it occurs.

Another form of sympatric speciation occurs when cytogenetic changes take place that result in "instantaneous speciation." These cytogenetic changes include polyploidy and translocations. For example, if polyploid offspring cannot produce fertile hybrids with individuals from a parent population, then the polyploid is reproductively isolated. This mechanism is much more common in plants because they can exist vegetatively despite odd ploidy and they usually do not have chromosomal sex-determining mechanisms, which are especially vulnerable to ploidy problems (see chapter 8).

The end result of cladogenesis is the divergence of a homogeneous population into two or more species. One of the classic examples of cladogenesis appears in

**Figure 21.4** The apple maggot fly, *Rhagoletis pomonella.* This species has exhibited host range expansion since the nineteenth century from hawthorn to apple, cherry, and roses. Host races are presumably the initial step in sympatric speciation. Magnification 10×. (Source: Jeffrey L. Feder and Guy L. Bush, Zoology Department, Michigan State University.)

the ground finches of the Galápagos Islands. These birds are very well studied not only because they present a striking case of speciation, but also because Darwin studied them and was strongly influenced by them in his views. Figure 21.5 is a map of the Galápagos Islands, and figure 21.6 is a diagram of the species of Darwin's finches.

An original flock of finches somehow reached the Galápagos Archipelago from South America, 700 miles away, and with time spread to the various islands of the Galápagos Archipelago. Given the limited ability of the birds to get from island to island, allopatric speciation took place. On each island, the finch population evolved reproductive isolating mechanisms while evolving to fill certain niches not already filled on the islands. For example, in South America, no finches have evolved to be like woodpeckers because many woodpecker species already live there. But the Galápagos Islands, being isolated from South America, have what is called a **depauperate fauna,** a fauna lacking many species found on the mainland. The islands lacked woodpeckers, and a very useful food resource for birds—insects beneath the bark of trees—was going unused. Finches that could make use of this resource would be at an advantage and would thus be favored by natural selection. On one island, a finch did evolve to use this food resource. The woodpecker finch acts like a woodpecker by inserting cactus needles into holes in dead trees to extract insects. Darwin wrote: "Seeing this gradation and diversity of structure in one small, intimately related group of birds, one might really fancy that from an original paucity of birds in this archipelago, one species had been taken and modified for different ends."

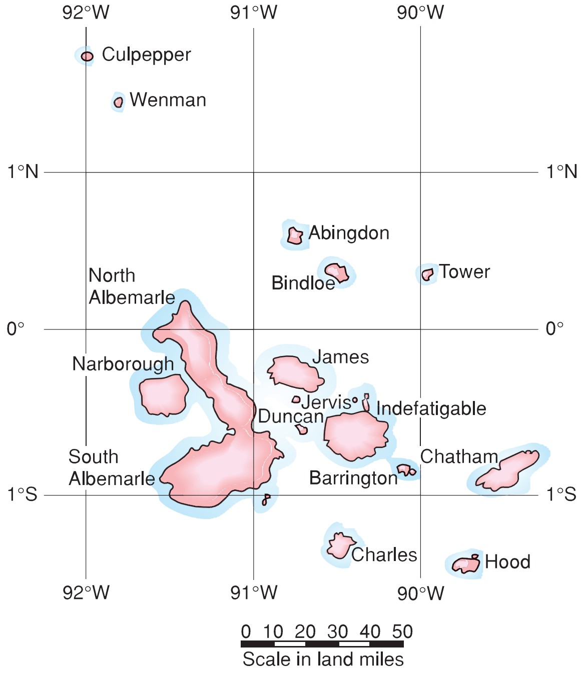
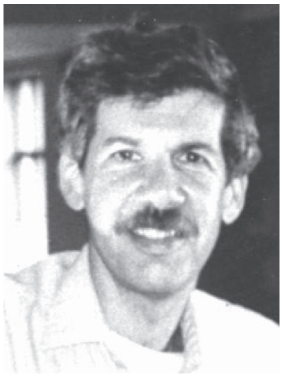


**Figure 21.5** The Galápagos Archipelago is located about 700 miles west of Ecuador. This isolated chain of islands is a natural laboratory for the study of evolutionary processes.
(From David Lack, *Darwin's Finches.* Copyright © 1947 by Cambridge University Press, New York, NY. Reprinted by permission.)

## Phyletic Gradualism Versus Punctuated Equilibrium

Darwin visualized cladogenesis as a gradual process, which we refer to as **phyletic gradualism.** However, an alternative view arose in 1972, when N. Eldredge and S. J. Gould suggested that speciation itself, and the morphological changes accompanying speciation, occur rapidly, separated by long periods of time when little change occurs (*stasis*). They called their model **punctuated equilibrium** (periods of stasis punctuated by rapid evolutionary change). Although figure 21.7 presents what appear to be two clear alternatives, in practice the models are very hard to tell apart. They both start with the same ancestral species and predict the same number of modern species. Allopatric, parapatric, and sympatric speciation mechanisms apply to both punctuated equilibrium and phyletic gradualism. The only major difference between the models is the rate of change, and this can only be discovered from an almost complete fossil record. The punctuated equilibrium model has brought much excitement to modern evolutionary biology. We await a time in the near future when we can decide which model has predominated in evolutionary history.



Stephen J. Gould (1943– ).
(Courtesy of Dr. Stephen J. Gould and the Harvard University News Office.)

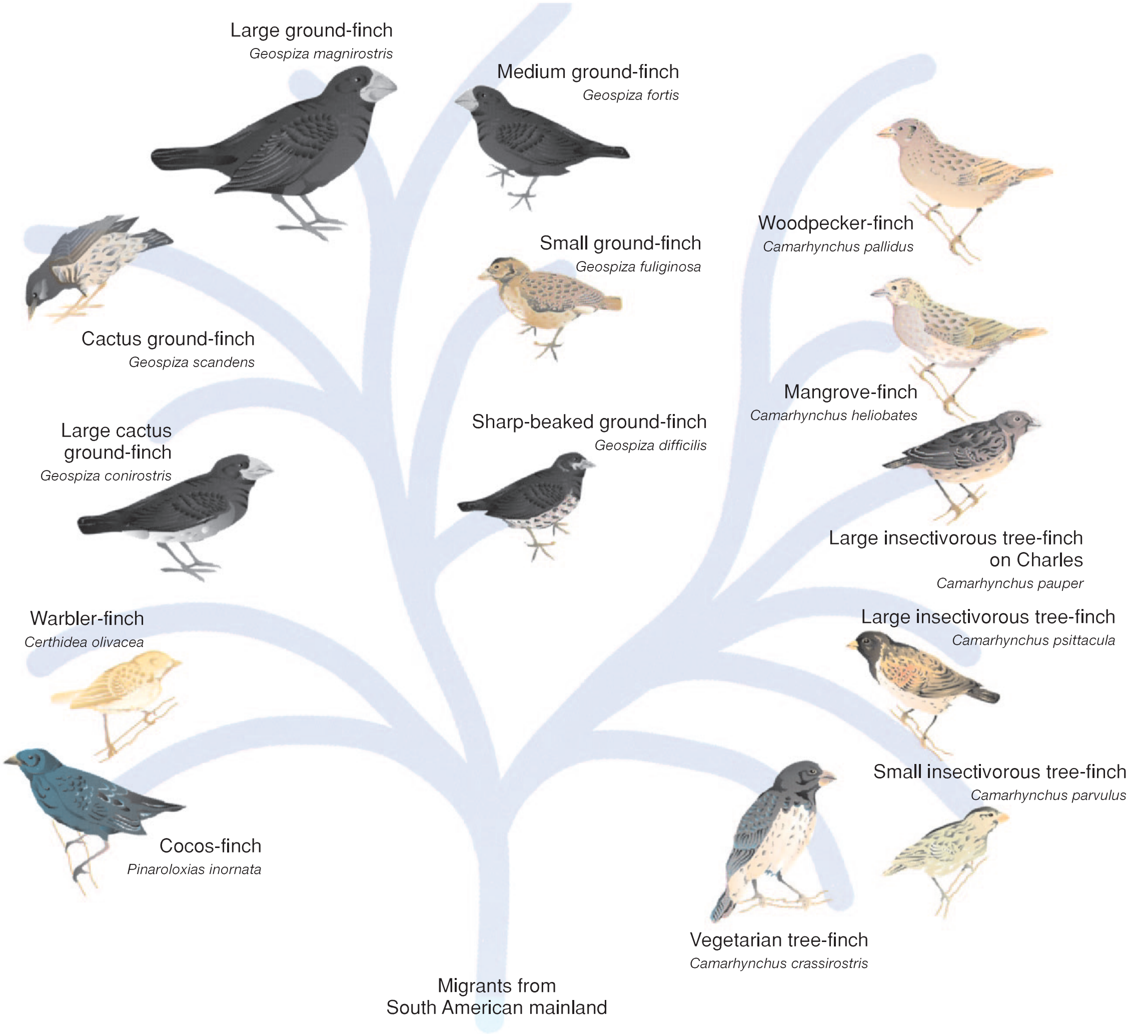

Large ground-finch
*Geospiza magnirostris*
Medium ground-finch
*Geospiza fortis*
Woodpecker-finch
*Camarhynchus pallidus*
Small ground-finch
*Geospiza fuliginosa*
Cactus ground-finch
*Geospiza scandens*
Mangrove-finch
*Camarhynchus heliobates*
Sharp-beaked ground-finch
*Geospiza difficilis*
Large cactus ground-finch
*Geospiza conirostris*
Large insectivorous tree-finch on Charles
*Camarhynchus pauper*
Warbler-finch
*Certhidea olivacea*
Large insectivorous tree-finch
*Camarhynchus psittacula*
Small insectivorous tree-finch
*Camarhynchus parvulus*
Cocos-finch
*Pinaroloxias inornata*
Vegetarian tree-finch
*Camarhynchus crassirostris*
Migrants from South American mainland

**Figure 21.6** Species of Darwin's finches. These birds apparently evolved from a single group of migrants from the South American mainland. Isolated on the different islands, the birds evolved to fill many vacant niches.

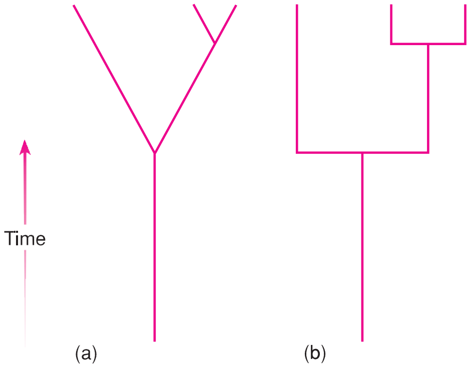


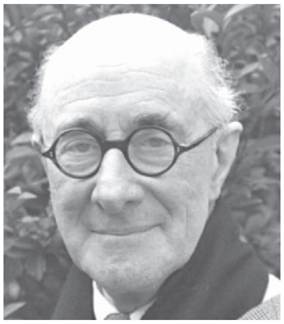
**Figure 21.7** Diagrammatic interpretation of cladogenesis. (*a*) Phyletic gradualism is depicted as a gradual divergence over time. (*b*) Punctuated equilibrium is depicted as a rapid divergence of two groups after long periods of no change. The horizontal axis is some arbitrary measure of species differences.

## GENETIC VARIATION

Darwinian evolution depends on the variation within a population. E. B. Ford, a British evolutionary biologist, applied the term **genetic polymorphism** to the occurrence of more than one allele at a given locus. Usually, we consider a locus polymorphic if a second allele occurs in the population at a frequency of 5% or more. Before the mid-1960s, the general belief was that only a few loci were polymorphic in any individual or any population.

In 1966, two researchers found a way to sample the genome in what they perceived was a random manner. R. C. Lewontin and J. L. Hubby used acrylamide-gel electrophoresis (see chapter 5) to investigate variability in a fruit fly species, *Drosophila pseudoobscura.* (H. Harris reported independent, similar work with human DNA.) Lewontin and Hubby reasoned that choosing enzymes and general proteins that are amenable to separation by electrophoresis, is, in fact, choosing a random sample of the genome of the fruit fly. If this is the case, then the degree of polymorphism found by electrophoretic sampling would provide an estimate of the amount of variability occurring in the individual organism and in the population. Their results were startling.

Lewontin and Hubby found that the species was polymorphic at 39% of eighteen loci examined, the average population was polymorphic at 30% of its loci, and the average individual was heterozygous at 12% of its loci. The high rate of polymorphism sparked two interrelated controversies. The first was whether electrophoresis does, in fact, randomly sample the genome. The second was whether most electrophoretic alleles are maintained in the population by natural selection. Let us return to the arguments after looking at ways in which genetic polymorphisms could be maintained in natural populations.



Edmund Brisco Ford (1901–1988). (Courtesy of Professor Edmund Brisco Ford.)

Richard C. Lewontin (1929– ). (Courtesy of Dr. Richard C. Lewontin.)

## Maintaining Polymorphisms

### Heterozygote Advantage

When selection acts against both homozygotes, an equilibrium is achieved, dependent solely on the selection coefficients, that maintains both alleles (see chapter 20). The classic example of heterozygote advantage in human beings is sickle-cell anemia. Sickle-cell hemoglobin ($Hb^S$) differs from normal hemoglobin ($Hb^A$) because it has a valine in place of a glutamic acid in position number 6 of the beta chain of the globin molecule. When the availability of oxygen is reduced, the erythrocytes containing sickle-cell hemoglobin change from round to sickle-shaped cells (see fig. 2.28). There are two unfortunate consequences: (1) sickle-shaped cells are rapidly broken down, which causes anemia as well as hypertrophy of the bone marrow, and (2) the sickle cells clump, which blocks capillaries and produces local losses of blood flow that result in tissue damage.

This condition of reduced fitness would lead one to predict that the sickle-cell allele would be selected against in all populations and, therefore, would be rare. But this is not the case. The sickle-cell allele is common in many parts of Africa, India, and southern Asia. What could possibly maintain this detrimental allele? In the search for an answer to this question, biologists discovered that the distribution of the sickle-cell allele coincided well with the distribution of malaria. The following facts have now been uncovered. The sickle-cell homozygote ($Hb^SHb^S$) almost always dies of anemia. The sickle-cell heterozygote ($Hb^AHb^S$) is only slightly anemic and has resistance to malaria. The normal homozygote ($Hb^AHb^A$) is not anemic and has no resistance to malaria. Thus, in areas where malaria is common, the most fit genotype of the three appears to be the sickle-cell heterozygote, which has resistance to malaria and only a minor anemia.

This conclusion is supported by the changes in allelic frequencies that occur when a population from a malarial area moves to a nonmalarial area. Since the normal homozygote is no longer at risk for malaria, selection acts mainly on the sickle-cell homozygote and, to a slight extent, on the heterozygote. Table 21.1 shows data for African blacks versus African Americans. The African population is, of course, under malarial risk, whereas the American population is not. The sickle-cell hemoglobin allele ($Hb^S$) is reduced in frequency in African Americans.

Heterozygote advantage is an expensive mechanism for maintaining a polymorphism. Losses must occur in both homozygous groups in order for the polymorphism to exist. Thus, part of the reproductive output of a population is lost each generation to maintain each polymorphism under heterozygote advantage. In the case of sickle-cell anemia, this means a tragic loss of human life due to either anemia or malaria. (The loss of individuals

**Table 21.1** **Sickle-Cell Anemia Frequencies in African Blacks and African Americans**

| | Percentage of Homozygotes ($Hb^AHb^A$) | Percentage of Heterozygotes ($Hb^AHb^S$) | Frequency of $Hb^S$ ($q$) |
|---|---|---|---|
| African Blacks (Midcentral Africa) | 82 | 18 | 0.09 |
| African Americans | 92 | 8 | 0.04 |

to maintain genetic variation at a particular locus is called **genetic load.** In the sickle-cell case, it is due to the segregation of individuals with lowered fitness and is therefore called **segregational load.**) Very few other examples of heterozygote advantage have been documented.

### *Frequency-Dependent Selection*

All the selection models discussed so far (chapter 20) have had selection coefficients that were constants. This is not always the case. For example, L. Ehrman has shown that when a female fruit fly has a choice between mates with different genotypes, the female fly chooses to mate with a male with a rare genotype. **Frequency-dependent selection** is selection in which the fitnesses of genotypes change according to their frequencies in the population.

The population geneticist Bruce Wallace has coined the terms *hard selection* and *soft selection* to deal with cases of frequency and density dependence. (Density-dependent selection exists when the fitness of a genotype changes as population density changes. We will not deal with that here.) Wallace defined soft selection as selection in which the selection coefficients depend on the frequency and density of genotypes. Hard selection is selection that is independent of both frequency and density. For example, the low fitness of sickle-cell anemia homozygotes involves hard selection because of the objectively deleterious effects of the anemia. Soft selection could be envisioned as selection that might act on aggressive behavioral genotypes in some lemming and field mouse species. When population density and frequency of the genotypes are low, these animals survive and reproduce. As population density increases, there can be a selection for more aggressive genotypes because they may be more successful in obtaining resources. As density increases further and the frequencies of the aggressive genotypes increase, they may be selected against because of the preoccupation of these aggressive individuals with territory defense under crowded conditions. This has been suggested as a mechanism of wildlife's "lemming cycle," rapid declines in the density of lemming and field mouse populations every three to five years.

Lee Ehrman (1935– ). (Courtesy of Dr. Lee Ehrman. Photo by Jan Robert Factor.)

A model for frequency-dependent selection can be constructed by assigning fitnesses that are not constants. One way to do this is to assign fitnesses that are a function of allelic frequencies. Thus, the assigned fitnesses for one locus with two alleles could be $(1.5 - p)$, 1, and $(1.5 - q)$ for the $AA$, $Aa$, and $aa$ genotypes, respectively (table 21.2). An interesting outcome of this model is that at $p = q = 0.5$, the system is in equilibrium, and no selection takes place because all the fitnesses are equal to 1.

Another way of looking at frequency-dependent selection is to look at the situation in which each genotype exploits a slightly different resource. As a genotype becomes rare, competition for the resource that genotype uses will likely decrease, and the genotype will thus have an advantage over the common genotypes, which are competing for resources. This type of selection is probably very common.

### *Transient Polymorphism*

A genetic polymorphism can result when an allele is being eliminated either by random or selective mechanisms. If a population starts out homozygous for the $a$ allele, for example, and a mutation brings in a more favored $A$ allele, the population gradually becomes all $A$ through directional selection. However, during the process of replacement, both alleles are present.

### *Other Systems*

Selection at one stage in the life cycle of an organism can balance a different form of selection at another stage in the life cycle. For example, an allele can be favored in a

**Table 21.2** Selection Model of Frequency-Dependent Selection: The *A* Locus with the *A* and *a* Alleles

| | Genotype | | | |
|---|---|---|---|---|
| | ***AA*** | ***Aa*** | ***aa*** | **Total** |
| Initial genotypic frequencies | $p^2$ | $2pq$ | $q^2$ | 1 |
| Fitness ($W$) | $1.5 - p$ | 1 | $1.5 - q$ | |
| Ratio after selection | $p^2(1.5 - p)$ | $2pq$ | $q^2(1.5 - q)$ | $\overline{W} = 0.5 + 2pq$ |
| Genotypic frequencies after selection | $\frac{p^2(1.5 - p)}{\overline{W}}$ | $\frac{2pq}{\overline{W}}$ | $\frac{q^2(1.5 - q)}{\overline{W}}$ | 1 |

larva but selected against in an adult. There can also be a balance of selection in different parts of the habitat in a heterogeneous environment. For instance, an allele can be favored in a wet part of the habitat but selected against in a dry part.

## Maintaining Many Polymorphisms

In summary, allelic polymorphisms in a population were classically accounted for by heterozygote advantage, frequency-dependent selection, or, infrequently, some other mechanism. Until Lewontin and Hubby did their work, heterozygote advantage was considered the most common method of maintaining a polymorphism at a given locus. The maintenance of an allele by heterozygote advantage costs the population a certain number of its offspring due to the mortality (or sterility) of the homozygotes. Most populations can afford the loss if polymorphisms are maintained at only a few loci. After Lewontin and Hubby reported that polymorphisms seemed to exist at a large proportion of loci, new explanations were needed to account for them. Three explanations were considered:

1. Electrophoresis (the technique used in Lewontin and Hubby's research) does not randomly sample the genome, and thus the large amount of variability they found does not really exist.
2. New population genetic models can be derived that explain how natural selection maintains this large amount of variability.
3. Electrophoretic alleles are not under selective pressure. That is, allozymic forms of an enzyme all perform the function of the enzyme equally well. This idea is called the **neutral gene hypothesis.**

### Sampling the Genome

Does electrophoresis randomly sample the genome? Since, on the basis of DNA content, the genome of higher organisms has the potential to contain half a million genes, this question may be difficult to resolve. Since the original reports of Lewontin and Hubby and Harris, numerous studies on many different organisms agree, for the most part, on the high amount of polymorphism in natural populations (table 21.3). However, several lines of evidence suggest that the results from electrophoresis are actually underestimates of the true amount of genetic variability present in a population.

The majority of amino acid substitutions, for example, do not change the charge of the protein. Thus, what appear to be single bands on an electrophoretic gel could actually be heterogeneous mixtures of the products of several alleles. Also, we now know that glycolytic enzymes are less polymorphic than other enzymes. Since glycolysis is a limited process in which most enzymes are not involved, it follows that the average heterozygosity over all loci should be slightly higher than the original estimates that included glycolytic enzymes. Recent technical advances of multidimensional electrophoresis and DNA sequencing support the hypothesis that electrophoresis does randomly sample the genome. However, DNA sequencing studies have shown that abundant variation exists, especially in the third (wobble) position of codons, and in parts of introns. Heterozygosity at the DNA sequence level seems to approach 100%.

### Multilocus Selection Models

Can standard genetic models account for the high degree of variability in natural populations? If each locus is considered independently, then for each polymorphic locus, offspring in a population lost to maintain that polymorphism by heterozygote advantage are independent of offspring lost due to selection at other loci. The losses would soon outstrip the reproductive capacity of any species. Models proposed since Lewontin and Hubby's report have suggested that natural selection favors the individuals that are the most heterozygous overall. Individuals selected against because of their homozygosity would be individuals with many homozygous loci. In other words, natural selection acts on the entire genome, not on each locus separately. We can show algebraically that the large

**Table 21.3** Survey of Genic Heterozygosity

| Species | Number of Populations | Number of Loci | Proportion of Loci Polymorphic per Population | Heterozygosity per Locus | Standard Error of Heterozygosity |
|---|---|---|---|---|---|
| *Homo sapiens* | 1 | 71 | 0.28 | 0.067 | 0.018 |
| *Mus musculus musculus* | 4 | 41 | 0.29 | 0.091 | 0.023 |
| *M. m. brevirostris* | 1 | 40 | 0.30 | 0.110 | — |
| *M. m. domesticus* | 2 | 41 | 0.20 | 0.056 | 0.022 |
| *Peromyscus polionotus* | 7 (regions) | 32 | 0.23 | 0.057 | 0.014 |
| *Drosophila pseudoobscura* | 10 | 24 | 0.43 | 0.128 | 0.041 |
| *D. persimilis* | 1 | 24 | 0.25 | 0.106 | 0.040 |
| *D. obscura* | 3 (regions) | 30 | 0.53 | 0.108 | 0.030 |
| *D. subobscura* | 6 | 31 | 0.47 | 0.076 | 0.024 |
| *D. willistoni* | 2–21 | 28 | 0.86 | 0.184 | 0.032 |
|  | 10 | 20 | 0.81 | 0.175 | 0.039 |
| *D. melanogaster* | 1 | 19 | 0.42 | 0.119 | 0.037 |
| *D. simulans* | 1 | 18 | 0.61 | 0.160 | 0.052 |
| *Limulus polyphemus* | 4 | 25 | 0.25 | 0.061 | 0.024 |

*Source: The Genetic Basis of Evolutionary Change* by R. C. Lewontin, (New York: Columbia University Press, 1974). Reprinted with permission of the publisher.

*Note:* See source (Lewontin, 1974) for individual references.

number of polymorphisms that exist in natural populations could be maintained according to these models.

### Neutral Alleles

The high incidence of polymorphism that electrophoresis reveals may not be important from an evolutionary point of view. If all or most electrophoretic alleles are neutral (i.e., if no allele is more fit than its alternative) or only very slightly deleterious, there is virtually no selection at these loci, and the variation observed in the population is merely a chance accumulation of a combination of mutation and genetic drift. This model, proposed by M. Kimura of Japan, is an alternative to the natural selection model.

Motoo Kimura (1924–1994). (Courtesy of Dr. Motoo Kimura.)

## Which Hypothesis Is Correct?

Researchers who favor the concept that most electrophoretic alleles are neutral do not deny that selection exists. They do not hold that evolution is non-adaptive, but say merely that most of the molecular variation (electrophoretic) found in nature is not related to fitness—it is neutral. Thus, the demonstration that selection actually exists, in electrophoretic systems or otherwise, is not proof against the neutralist view. No one denies the explanation for the maintenance of sickle-cell anemia. Selection at several other electrophoretic systems is also known.

For example, R. Koehn showed that different alleles of an esterase locus in a freshwater fish in Colorado produced proteins with different enzyme activities at different water temperatures. Koehn then showed that the alleles were distributed as one would predict on the basis of the water temperature. In other words, the distribution of alleles correlated with the distribution of water temperature. The enzyme produced by the $ES\text{-}1^a$ allele functioned best at warm temperatures, whereas the enzyme produced by the $ES\text{-}1^b$ allele functioned best at cold temperatures. The cold-adapted enzyme was prevalent in the fish in colder waters (higher latitudes), and the warm-adapted enzyme was prevalent in the fish in warmer waters (lower latitudes; fig. 21.8).

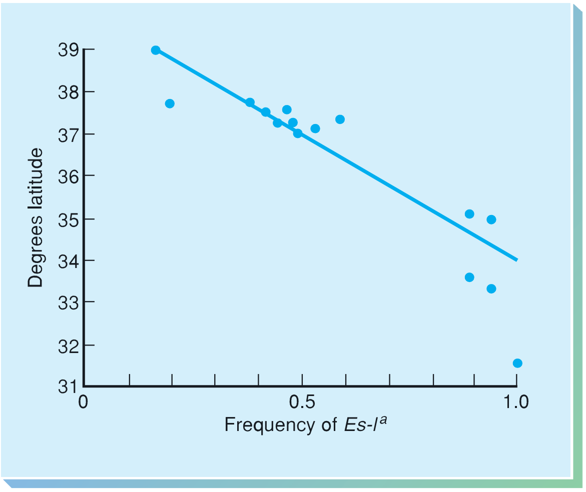


**Figure 21.8** Relation of latitude and frequency of the warm-adapted esterase allele *Es-I*$^a$ in populations of the fish *Catostomus clarki.* Note how the frequency of the allele increases as latitude decreases (warmer water). (From Richard Koehn, "Functional and evolutionary dynamics of polymorphic esterases in catostomid fishes," *Transactions of the American Fisheries Society,* 99:223. Copyright © 1970 American Fisheries Society, Bethesda, MD.)

Isolated instances of selection, however, do not adequately prove the case for maintaining variation by means of natural selection or disprove the case for maintaining variation of neutral alleles. Both theories recognize natural selection as the guiding force in producing adapted organisms. What is needed is proof that the majority of polymorphic loci are either being selected or are neutral. For this proof, many loci must be examined independently—a very difficult undertaking—or some grand pattern must emerge supporting one hypothesis or the other.

## Grand Patterns of Variation

### Clinal Selection

Data on the geographic distribution of alleles fail to adequately support either theory. Often, a single allele predominates over the range of a species (fig. 21.9). Changes in allelic frequency from one geographic area to another can often be attributed to **clinal selection,** selection along a geographic gradient, in which allelic frequencies change as altitude, latitude, or some other geographic attribute changes. Note in figure 21.9 the general increase in the *Es-5*$^b$ frequency from west to east in the southern United States. But, in line with the neutralist view, geographic patterns similar to those in figure 21.9 can also be produced by neutral alleles with a very low level of migration, as little as one individual per one thousand per generation.

### Molecular Evolutionary Clock

The advancing technology that made it possible to detect the sequence of amino acids in a protein also made it possible to discover how much the proteins and DNA of various species differ. In chapter 17, we discussed the use of mitochondrial DNA (mtDNA) to determine evolutionary relationships. Currently, protein, nuclear DNA, and mtDNA clocks are being studied.

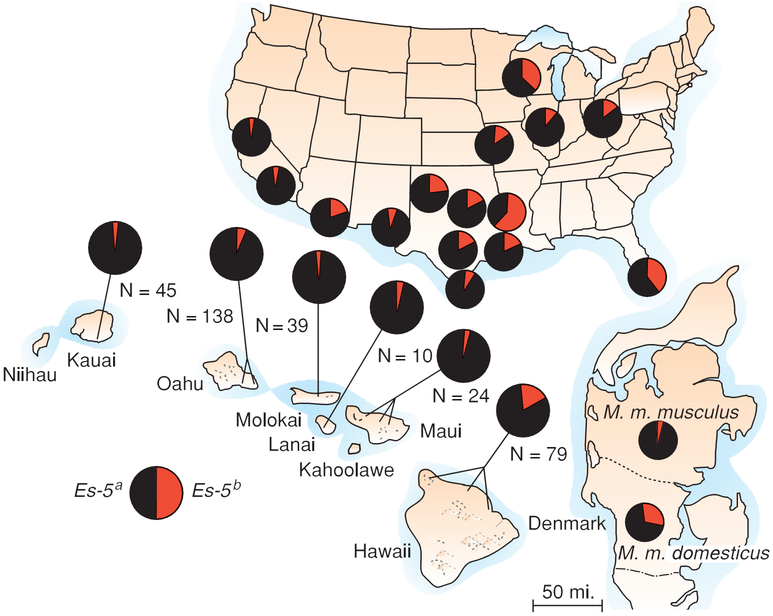


**Figure 21.9** Frequency distribution of the *Es-5* alleles of an esterase locus in house mice. Each *circle* represents allelic frequencies at that geographic location. Note the general tendency for the *Es-5*$^b$ allele to increase from west to east in the continental United States. (From L. Wheeler and R. Selander, "Genetic Variation in Populations of the House Mouse, Mus musculus, in the Hawaiian Islands," *Studies in Genetics,* VII, 1972. University of Texas Publication 7213. Reprinted with permission of M. R. Wheeler.)

Knowledge of the changes in amino acid sequences can be used to estimate the rate of evolutionary change. That is, the data show how many amino acid substitutions have occurred between two known groups of organisms. The genetic code dictionary allows us to estimate the minimum number of nucleotide substitutions required for this change. For example, if one protein contains a phenylalanine in position 7 (codons UUU, UUC), and the same protein in a different species has an isoleucine in the same position (AUU, AUC, AUA), we can see that the minimum number of substitutions to convert a phenylalanine codon to an isoleucine codon is one (UUU → AUU). When we know the minimum number of substitutions, we can calculate molecular **evolutionary rates,** nucleotide substitutions per million years. In a sense, these rates provide us with a **molecular evolutionary clock** that measures evolutionary time in nucleotide substitutions.

Many studies of the rate of amino acid and nucleotide substitutions have been done on hemoglobin, on cytochrome *c*, on a class of proteins involved in blood clotting called fibrinopeptides, and on many others. Figure 21.10 shows the way in which an amino acid sequence differs among species. From comparisons of this type, we can calculate the actual number of amino acid differences, as well as percentage differences. Table 21.4 is a compilation of percentage differences between various species based on the cytochrome *c* protein. This type of information can be used two ways.

First, we can construct a **phylogenetic tree** that tells us the evolutionary history of the species under consideration (fig. 21.11). This tree can be compared with phylogenetic trees constructed by more classical means using fossil evidence and evidence from morphology, physiology, and development. From the comparisons, we can look at areas of disagreement in an attempt to find out the best way to create phylogenetic trees. In addition, molecular phylogenies can give us information unattainable in any other way, as, for example, when the fossil record is incomplete or ambiguous.

A second use of DNA or amino acid difference data is to determine average rates of substitution. Once we know the current amino acid differences in the proteins of two species, it is possible to estimate the actual number of nucleotide substitutions that have taken place over evolutionary time using the statistical Poisson distribution, which deals with rare events. The index, $K$, is the average number of amino acid substitutions, per site, between two proteins:

$$K = -\ln(1 - p)$$

in which ln is the natural logarithm (to the base $e$), and $p = d/n$ in which $d$ is the number of amino acid differences and $n$ is the total number of amino acid sites being compared. For example, in figure 21.10, $n = 8$ and $d = 3$ between the dog and chicken. Thus

$$K = -\ln(1 - 0.375) = 0.47$$

Therefore, the average number of amino acid substitutions, per site, between dog and chicken is 0.47.

We can take this calculation one step further by determining the per-year rate:

$$k = K/2T$$

in which $k$ is the amino acid substitution rate per site per year, and $T$ is the number of years since the two species diverged from a common ancestor. We divide by $2T$ because each side of the tree has evolved independently for $T$ years. When $k$'s are calculated for many proteins over many species, they cluster around $10^{-9}$ (table 21.5). In fact, Kimura has suggested the unit of a *pauling* to be equal to $10^{-9}$ amino acid substitutions per year per site in

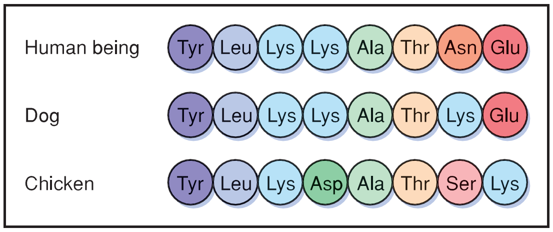


**Figure 21.10** The amino acids making up the terminal portion of cytochrome *c* in three species. Note the similarities and differences.

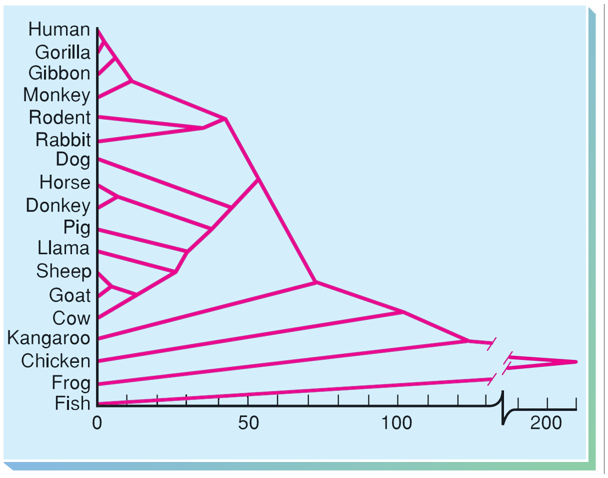


**Figure 21.11** Composite evolution of hemoglobin, cytochrome *c*, and fibrinopeptide A. The total number of nucleotide substitutions appears on the horizontal axis. Note how the tree groups similar organisms and generally agrees with classical systematics. (From C. H. Langley and W. M. Fitch, "An examination of the constancy of the rate of molecular evolution," *Journal of Molecular Evolution,* 3:168. Copyright © 1974 Springer-Verlag, Heidelberg. Reprinted by permission.)

**Table 21.4** **Amino Acid Differences (By Percentage) in Cytochrome *c* Between Different Organisms**

| | Human being | Pig | Horse | Chicken | Turtle | Bullfrog | Tuna | Carp | Lamprey | Fruit fly | Screwworm | Silkworm | Sesame | Sunflower | Wheat | *C. krusei* | Yeast | *N. crassa* | *R. rubrum* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Human being | 0 | 10 | 12 | 13 | 14 | 17 | 20 | 17 | 19 | 27 | 25 | 29 | 35 | 38 | 38 | 46 | 41 | 44 | 65 |
| Pig, bovine, sheep | | 0 | 3 | 9 | 9 | 11 | 16 | 11 | 13 | 22 | 20 | 25 | 38 | 40 | 40 | 45 | 41 | 43 | 64 |
| Horse | | | 0 | 11 | 11 | 13 | 18 | 13 | 15 | 22 | 20 | 27 | 39 | 41 | 41 | 46 | 42 | 43 | 64 |
| Chicken, turkey | | | | 0 | 8 | 11 | 16 | 14 | 17 | 23 | 21 | 26 | 40 | 41 | 41 | 45 | 41 | 44 | 64 |
| Snapping turtle | | | | | 0 | 10 | 17 | 13 | 18 | 22 | 22 | 26 | 38 | 39 | 41 | 47 | 44 | 45 | 64 |
| Bullfrog | | | | | | 0 | 14 | 13 | 20 | 20 | 20 | 27 | 41 | 42 | 43 | 46 | 43 | 45 | 65 |
| Tuna | | | | | | | 0 | 8 | 18 | 23 | 22 | 30 | 42 | 43 | 44 | 43 | 43 | 45 | 65 |
| Carp | | | | | | | | 0 | 12 | 21 | 20 | 25 | 40 | 41 | 42 | 45 | 42 | 43 | 64 |
| Lamprey | | | | | | | | | 0 | 27 | 26 | 30 | 44 | 44 | 46 | 50 | 45 | 47 | 66 |
| Fruit fly | | | | | | | | | | 0 | 2 | 14 | 42 | 41 | 42 | 43 | 42 | 38 | 65 |
| Screwworm fly | | | | | | | | | | | 0 | 13 | 41 | 40 | 40 | 43 | 42 | 38 | 64 |
| Silkworm moth | | | | | | | | | | | | 0 | 39 | 40 | 40 | 43 | 44 | 44 | 65 |
| Sesame | | | | | | | | | | | | | 0 | 10 | 13 | 47 | 44 | 48 | 65 |
| Sunflower | | | | | | | | | | | | | | 0 | 13 | 47 | 43 | 49 | 67 |
| Wheat | | | | | | | | | | | | | | | 0 | 45 | 42 | 48 | 66 |
| *Candida krusei* | | | | | | | | | | | | | | | | 0 | 25 | 39 | 72 |
| Baker's yeast | | | | | | | | | | | | | | | | | 0 | 38 | 69 |
| *Neurospora crassa* | | | | | | | | | | | | | | | | | | 0 | 69 |
| *Rhodospirillum rubrum* | | | | | | | | | | | | | | | | | | | 0 |

*Source:* From M. O. Dayhoff, ed., Atlas of Protein Sequence and Structure," National Biomedical Research Foundation, Washington, D.C., 1972. Reprinted with permission.

honor of Linus Pauling, who, along with E. Zuckerkandl, first proposed the concept of a molecular clock in 1963. If the values of *k* (such as those in table 21.5) form a normal distribution around $10^{-9}$, then $10^{-9}$ would be the rate of "the" molecular evolutionary clock. So far, the data have been too limited to determine the distribution.

Although controversy still exists, the neutralists have interpreted the relative constancy of the molecular evolutionary clock as strong evidence in support of the neutral gene hypothesis. A constant rate of molecular evolution over many groups of organisms over many different time intervals implies that the substitution rate is a stochastic or random process rather than a directed or selectional process. This is not to say that no adapted changes occur in proteins or that there are no constraints. In fact, the evidence suggests that three classes of amino acids can be grouped in terms of substitution rate: invariant, moderately variant, and hypervariant. It seems possible that virtually no substitutions of amino acids will occur in and around the active site of the enzyme since any amino acid change in that area might be deleterious or lethal. For example, a segment of cytochrome *c* that runs from amino acids 70 to 80 is invariant in all organisms tested. This area includes a binding site of the protein.

### *DNA Variation*

If the neutralist view of molecular evolution is correct, we should be able to make some predictions about rates of change in DNA. For example, we predict that DNA under greater constraint should amass fewer base changes than DNA under lesser constraint. We could test this by looking at the accumulation of mutations in the three positions of the codon, or we could look at DNA that is not directly translated, such as pseudogenes (see chapter 15) or introns, which are probably under lesser constraint. Let us first look at the three positions of the codon.

A reexamination of the codon dictionary (see table 11.4) shows that the third, or wobble, position of the codon should be under less constraint. Eight amino acids belong to unmixed families; their amino acids are defined by the first and second positions coupled with any of the

**Table 21.5** **Evolutionary Rates (*k*) as $10^{-9}$ Substitutions Per Amino Acid Site Per Year for Various Proteins**

| Protein | *k* |
|---|---|
| Fibrinopeptide | 8.3 |
| Pancreatic ribonuclease | 2.1 |
| Lysozyme | 2.0 |
| Hemoglobin alpha | 1.2 |
| Myoglobin | 0.89 |
| Insulin | 0.44 |
| Cytochrome *c* | 0.3 |
| Histone H4 | 0.01 |

*Source:* From M. Kimura, *The Neutral Theory of Molecular Evolution*, Cambridge University Press, 1983. Reprinted with the permission of Cambridge University Press.

four bases in the third position of the codon. The remaining amino acids belong to mixed families; the first two positions and the purine or pyrimidine nature of the third position in their codons is important. Hence, the wobble (third) position of the codon is under the least constraint and should build up the most neutral or near-neutral mutations.

In addition, analysis of changes in the first and second positions indicates that more drastic change takes place by mutation of the second rather than the first position of the codon. Thus, we predict that evolutionary distance, as measured by base substitutions, should be greatest for the third codon position and least for the second position. This turns out to be generally true (table 21.6).

It should be clear that a major problem facing those who study evolutionary clocks is how to calibrate them. Are average changes uniform throughout lineages? Do clocks speed up, slow down, or show other unpredictable changes through time? There is evidence, for example, that both the nuclear and mitochondrial DNA clocks have slowed down in the hominid lineage as compared with old world monkeys. If the clocks change speed in different lineages, at different times, and for different parts of the genome, there will be errors in interpreting lineages and errors in using averages to understand the general patterns of change.

**Table 21.6** **Evolutionary Distance of Codons, Measured in Base Substitutions Per Nucleotide Site**

| | Codon Site | | |
|---|---|---|---|
| | 2 | 1 | 3 |
| Beta globin, human being vs. mouse | 0.13 | 0.17 | 0.34 |
| Beta globin, chicken vs. rabbit | 0.19 | 0.30 | 0.64 |
| Rabbit, alpha vs. beta globin | 0.44 | 0.54 | 0.90 |

*Source:* From M. Kimura, *The Neutral Theory of Molecular Evolution*, Cambridge University Press, 1983. Reprinted with the permission of Cambridge University Press.

At this point, it is probably safe to say that while natural selection acts to create organisms that are adapted to their environments (see box 21.2 on mimicry), many nucleotide and amino acid changes may not have measurable effects on the fitness of the organism, and hence their frequencies may be determined by the stochastic processes of mutation and genetic drift. Adaptation is by natural selection, but neutral variation most certainly also occurs in organisms.

## SOCIOBIOLOGY

We close this chapter by looking at a level of evolution only recently addressed. In 1975, E. O. Wilson published a mammoth tome entitled *Sociobiology: The New Synthesis.* This book has been at the center of major controversies that have spread to the fields of sociology, psychology, anthropology, ethology, and political science. The basic premise of the book is that social behavior is under genetic control. Although Wilson's book contains twenty-six chapters concerned with the animal kingdom, controversies have arisen because of the one chapter that applies the theory to human beings.

### Altruism

V. C. Wynne-Edwards published a book in 1962 entitled *Animal Dispersion in Relation to Social Behavior.* In it, he suggested that animals regulate their own population density through altruistic behavior. For example, under crowded conditions, many birds cease reproducing. The interpretation of this phenomenon was that these birds were being altruistic: Their failure to breed was for the ultimate good of the species. (**Altruism** means risking loss of fitness in an act that could improve the fitness of another individual.) Wynne-Edwards suggested a mechanism called **group**

Edward O.Wilson (1929– ). (Courtesy of Dr. Edward O. Wilson. Photo by Pat Hill/OMNI Publications International, Ltd.)

## BOX 21.2

### Experimental Methods

### *Mimicry*

**Mimicry** is a phenomenon whereby an individual of one species gains an advantage by resembling an individual of a different species. There are at least two types of mimicry.

In **Mülerian mimicry,** named after F. Müller, several groups of organisms gain an advantage by looking like one another. This mimicry occurs among organisms in which all the mimetic species are offensive and obnoxious. The classic example is the general similarity among bees, wasps, and hornets.

In **Batesian mimicry,** named after H. W. Bates, a vulnerable organism (mimic) gains a selective advantage by looking like a dangerous or distasteful organism (model). The classic example of Batesian mimicry was, until 1991, the monarch (*Danaus plexippus*) and viceroy (*Limenitis archippus*) butterflies (fig. 1). Although the viceroy is smaller and, on close examination, looks different from the monarch, the resemblance is striking at first glance. Monarch butterflies feed on milkweed plants, obtaining noxious chemicals called cardiac glycosides, which the monarchs store in their bodies. When a bird tries to eat a monarch, it becomes sick and regurgitates what it has eaten. Thereafter, the bird will not only avoid eating monarchs, but it will also avoid eating any butterflies that look anything like monarchs. Previously it was believed that the mimetic viceroy butterfly gained a selective advantage by looking like the monarch and fooling bird predators into thinking that the viceroy was bad to eat. However, D. Ritland and L. Brower demonstrated a previously unrealized fact: The viceroys taste as bad as the monarchs to birds. This fact changes the mimicry of these two species from Batesian to Müllerian mimicry.

Examples of Batesian mimicry do occur in numerous butterfly species. For example, in West Africa, *Pseudacraea* species mimic species of the genus *Bematistes* (fig. 2). These species are primarily black and white or black and orange, and in some the sexes differ, each having a different mimic. Upwards of twenty species can be involved in these mimicry complexes in one area. Both forms of mimicry depend on the selective pressure generated by predation. Certain requirements must be met for each system to work properly. Batesian mimicry has the following requirements:

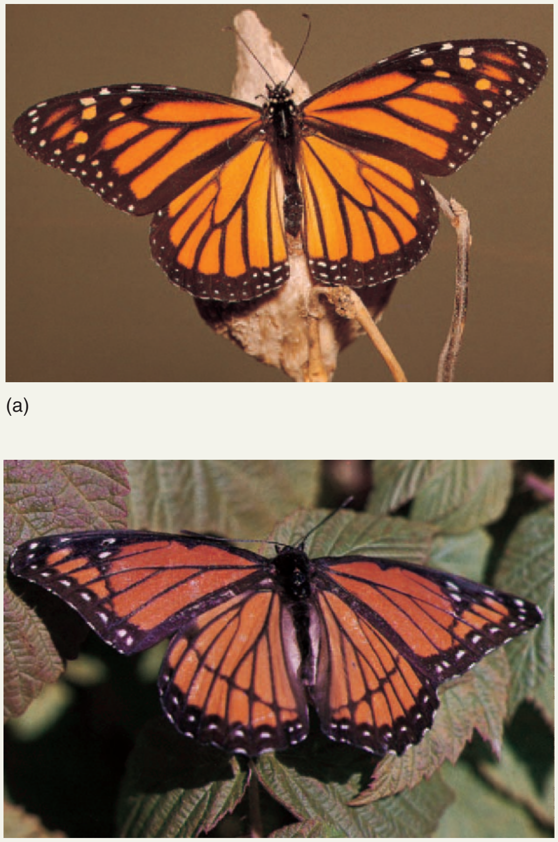
1. The model species must be conspicuous and inedible or dangerous.
2. Both model and mimic species must occur in the same area, with the model being very abundant. If the model is rare, predators do not have sufficient opportunity to learn that its pattern is associated with a bad taste. In fact, the reverse can happen; the model can be at a selective disadvantage if it is rare because the predators will learn from the mimic that the pattern is associated with something good to eat.
3. The mimic should be very similar to the model in the morphological characteristics predators perceive but not necessarily similar in other traits. The mimic is not evolving to *be* the model, only to look like it.

Müllerian mimicry requires that all the species be similar in appearance and distinctly colored. They can, however, be equally numerous. And, as the British geneticist P. M. Sheppard pointed out, the resemblance among Müllerian mimics need not be as good as between the mimic and model of a Batesian pair because Müllerian mimics are not trying to deceive a predator, only to remind the predator of the relationship.

Although there have been some critics of mimicry theory, especially critics of the way in which the system could evolve, the general model put forth by population geneticist and mathematician R. A. Fisher is generally accepted. According to Fisher, any new mutation that gave a mimic any slight advantage would be selected for. As time proceeded, other loci that might favorably modify the expression of mimetic genes would also be selected for in order to increase the similarity of mimic and model. This mechanism surmounts the criticism that a single mutation could not produce a mimic that so closely resembled its model.

**selection:** groups that had altruistic behavior would have a survival advantage over groups that did not.

In 1966, G. Williams, in his book *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought,* refuted the altruistic view with the charge that individuals that performed altruistic acts would be selected against. In other words, organisms not performing altruistic acts would have a higher degree of fitness. Williams held that apparent altruism had to be interpreted on the basis of benefits accruing to the individual performing the altruistic act. After his book, the idea of doing something for the good of the species became passé. How, then, can apparent altruism be accounted for? How can we explain why ground squirrels appear to put themselves at risk to predators by giving alarm calls, and why female workers in ant, wasp, and bee colonies



(a)

(b)

**Figure 1** Müllerian mimicry. (*a*) Monarch butterfly and (*b*) viceroy butterfly. Both have similar colors (*orange* and *black*) and a generally similar color pattern. ([*a*] © Robert Finke/Photo Researchers, Inc. [*b*] © Richard Parker/Photo Researchers, Inc.)

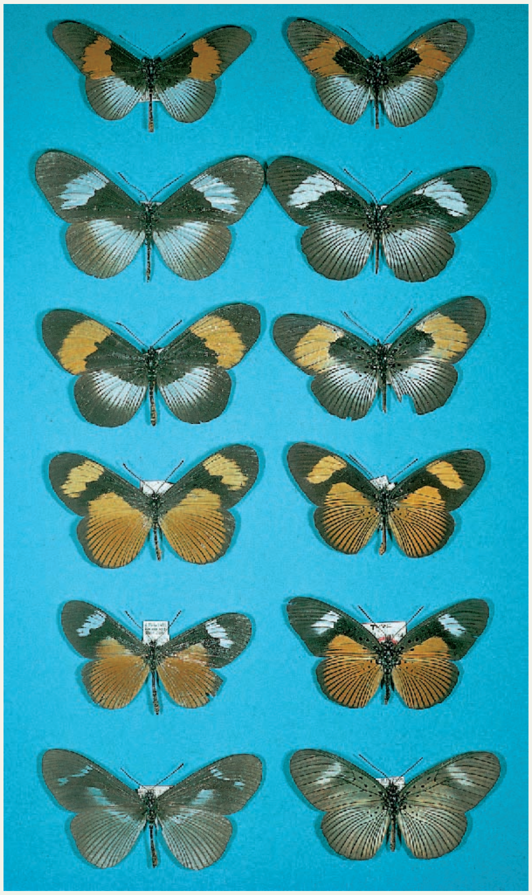


**Figure 2** Batesian mimicry seen in West African butterflies that live in the same places. Those on the *left* are species belonging to the genus *Bematistes*. Those on the *right* that mimic them are different species belonging to the genus *Pseudacraea*. (© J. A. L. Cooke/Oxford Scientific Films/Animals, Animals.)

forsake reproduction in order to work for the colony? **Sociobiology,** the study of the evolution of social behavior, attempts to answer these questions.

## *Kin Selection and Inclusive Fitness*

In 1964, W. D. Hamilton developed concepts that explained altruistic acts without resorting to group selection. Starting with the known fact that relatives have alleles in common, Hamilton suggested that natural selection would favor an allele that promoted altruistic behavior toward relatives because the result might be an increase in copies of that allele in the next generation. The proportion of alleles shared by two individuals can be defined as a **coefficient of relationship, *r*.** If an individual has a certain allele, the probability that a particular relative also has that allele is *r.* Siblings have an $r = 1/2$. A squirrel is likely to have virtually all its alleles still viable if it sacrifices itself for two or more siblings. In fact, natural selection should definitely favor altruism of an individual toward three siblings because, in a sense, natural selection is weighing 1 copy of an individual's alleles (the individual itself) versus 1.5 copies (three siblings).

This sort of reasoning has been termed the **calculus of the genes.** It does not imply that individuals actually think these things out; rather, natural selection has favored the individuals that behave this way. Hamilton referred to the sum of an individual's fitness plus the fitness effects of alleles that relatives share as **inclusive fitness.** He referred to the way natural selection acts on inclusive fitness as **kin selection.**

Hamilton applied his ideas of inclusive fitness and kin selection to explain sterile castes in the eusocial (truly social) hymenoptera (bees, ants, and wasps). The workers in these colonies are sterile females. Why do they forsake their ability to reproduce in order to help maintain the hive or colony? The answer seems to come from **haplodiploidy,** the unusual sex-determining mechanism of these species. In the eusocial hymenoptera with sterile castes, fertilized eggs produce diploid females, whereas unfertilized eggs produce haploid males (drones). The difference between a reproductive queen and a sterile worker in bees is larval nutrition: larvae fed "royal jelly" can become queens. Hamilton showed that since a worker is more closely related to her sisters than to her own potential offspring, kin selection could favor a worker who helps her sisters at the expense of her own reproduction.

Figure 21.12 shows a queen (female) with alleles $A_1$ and $A_2$ at the $A$ locus and a haploid drone (male) with the $A_3$ allele. A daughter will have either the $A_1A_3$ or $A_2A_3$ genotype. If we compare one of these daughters with her sisters, we see that the average $r = 0.75$—half of the time, $r = 1.0$, and the other half of the time, $r = 0.5$. A queen and her daughters have an $r = 0.5$. Thus, we see that workers (females) are more closely related to their sisters, and hence are at a reproductive advantage by raising them rather than their own young. Wilson has pointed out that sterile caste systems have evolved among insects in only one other group beside the eusocial hymenoptera, the termites. Although eusocial hymenoptera make up only 6% of insects, sterile castes have independently evolved at least eleven times. This is compelling evidence for the validity of Hamilton's analysis. Only one noninsect example of a caste has been discovered: the naked mole rat, a small subterranean rodent living in Africa, has this type of social system.

Many studies concerned with apparently altruistic acts have provided a large body of support for Hamilton's theory of kin selection and inclusive fitness. P. Sherman, working with ground squirrels, for example, has observed that the individuals that make the alarm calls have the most to gain from the standpoint of inclusive fitness; these individuals are resident females surrounded by kin.

One other explanation for altruism is also consistent with benefits to individual fitness. It is that many apparently altruistic acts are in reality selfish—they just look altruistic. To be altruistic, an individual must risk reducing its fitness to potentially benefit the fitness of others. We may, in fact, misinterpret some acts as altruistic that simply are not.

This turnaround in thought, from group selection to individual selection, has been an intellectual revolution in modern evolutionary biology. Before this revolution, many of the behaviors in nature that involved apparent altruism were difficult to explain. Now sociobiological reasoning provides an explanation.

The reason so much controversy has sprung up over the theory of genetic control of social behavior is because of the implications the theory has for human social, political, and legal issues. Human husband-wife, parent-child, and child-child conflicts, for example, may be built into the genes. Altruism, our highest form of nobility, may be mere selfishness. Many critics fear that sociobiological concepts can be used to support sexism and racism. For human beings, the alternative to the theory of sociobiology is the theory that most human behavior, including cultural learning, is determined by the environment. At present, although much evidence remains to be gathered, the sociobiology concept is compelling to many evolutionists.

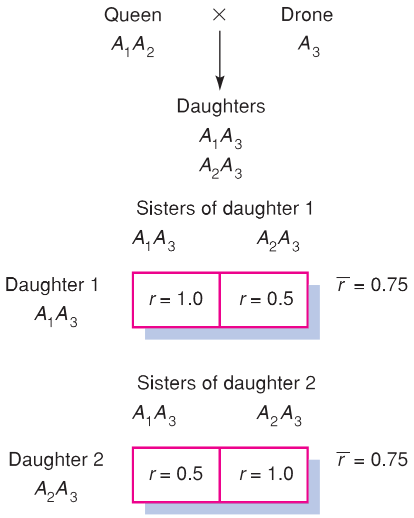


**Figure 21.12** Haplodiploidy in eusocial hymenoptera produces sisters with an average *r* of 0.75. Because drones (males) are haploid, queens produce daughters of only two genotypes at any locus. A given daughter has an *r* of 1.0 with sisters of identical genotype and an *r* of 0.5 with sisters of the other genotype, for an average *r* of 0.75. In other words, females have a 75% genetic similarity with their sisters but only a 50% similarity with their own offspring.

## SUMMARY

**STUDY OBJECTIVE 1:** To analyze the mechanisms of evolution and speciation 589–595

The theory of evolution by natural selection was put forward by Charles Darwin, who recognized the natural variation among individuals within a population of similar organisms. He noted also that offspring are overproduced in nature, and this overproduction inevitably leads to competition for scarce resources. Darwin assumed that, when competition occurs, the most fit will survive; through time, then, a population will become better adapted to its environment through the process of natural selection. Applying the algebra of population genetics to this theory leads to the modern concept of evolution, neo-Darwinism.

Cladogenic speciation occurs when reproductive isolating mechanisms arise, usually after gene flow in a population is blocked. Different populations of a species can then evolve independently. When individuals from the isolates can no longer interbreed, speciation has taken place. If the isolates then come in contact again, they will remain as separate species. Speciation may occur gradually or in a punctuated manner; it can be by allopatric, parapatric, or sympatric mechanisms.

**STUDY OBJECTIVE 2:** To investigate the mechanisms of the maintenance of genetic variation in natural populations, both selective and neutral 596–603

Evolution depends on variation. In 1966, Lewontin and Hubby, using electrophoresis, showed that a tremendous amount of heterozygosity occurred in natural populations. Attempts to explain this variation have led to two major competing theories: (1) variation is maintained selectively and (2) variation is not under selective pressure, but is instead neutral. Two areas of evidence support the neutralist view.

First, the molecular evolutionary clock (the per-year, per-amino acid, substitution rate) appears to be fairly constant at $10^{-9}$. This constancy implies that the majority of amino acid changes are the result of stochastic processes. Second, there have been greater numbers of nucleotide substitutions in DNA under lesser constraint than in DNA under greater constraint. For example, the third, or wobble, position of the codon has accumulated more mutations than the other two positions. We conclude that natural selection creates adapted organisms, but the majority of base and amino acid changes may be neutral.

**STUDY OBJECTIVE 3:** To discuss sociobiology, the evolution of social behavior 603–606

*Sociobiology* is another term for evolutionary behavioral ecology. It attempts to provide evolutionary explanations for social behaviors. Apparent altruistic behavior can be explained either as kin selection or as selfishness. Sterile insect castes have come about because of the unusual haplodiploid sex-determining mechanism in the eusocial hymenoptera. There is much controversy about and little information for applying sociobiological principles to human behavior.

## SOLVED PROBLEMS

**PROBLEM 1:** What are the roles of reproductive isolating mechanisms in the process of evolution?

***Answer:*** Reproductive isolating mechanisms prevent individuals in two populations from mating with each other or producing viable offspring. These mechanisms can be prezygotic or postzygotic. They usually evolve while populations are isolated from each other, either physically or during parapatric or sympatric speciation. For example, if a species is split by a new river, the populations on either side of the river can evolve in isolation from each other. Reproductive isolating mechanisms usually evolve irrespective of the other facets of evolution taking place. Thus, if, after time, the two populations come into contact (the river dries up), reproductive isolating mechanisms may have evolved to prevent mating. If weak reproductive isolating mechanisms have evolved, natural selection usually favors strengthening them by selecting against hybrids and against any mating behavior that leads to the formation of hybrids.

**PROBLEM 2:** What is our modern evolutionary concept of altruism?

***Answer:*** An altruistic act is one in which an individual risks the loss of fitness in order to benefit another individual. Human beings value these “selfless” acts; however, they are not favored in natural animal populations, except under very specific circumstances, because altruistic acts should be selected against. In other words, all other things being equal, an individual that did not do altruistic acts would have a higher fitness than one that did do these acts. Therefore, fitness is higher for "selfish" individuals. Altruistic acts, however, are expected if the beneficiary of the acts shares genes in common with the benefactor performing the acts. Generally, altruism can be expected among relatives, following the rules of kin selection.

## EXERCISES AND PROBLEMS*

### DARWINIAN EVOLUTION

**1.** Outline the Darwinian mechanism of the process of evolution. What is meant by neo-Darwinism?

### EVOLUTION AND SPECIATION

**2.** Population geneticist Hampton Carson has defined a "population flush" as a period of reduced selection during population increase. Why should there be reduced selection during a flush?

**3.** Describe how the processes of allopatric, parapatric, and sympatric speciation could take place.

**4.** Can information on evolutionary rates gained from molecular techniques shed light on the punctuated equilibrium-phyletic gradualism controversy? What additional data are needed to decide this controversy?

**5.** What is meant by "constraint" in the molecular evolution of DNA and proteins?

**6.** Recently, a vial of bull semen was stolen from an artificial insemination facility. Your friend is about to undergo artificial insemination and is concerned she may give birth to a Minotaur, or a cow-human hybrid. Provide two explanations for why she should not worry about this possibility.

**7.** In *Drosophila,* females in populations A and B produce an average of 250 offspring each. When the two populations are crossed, AB females produce only about 100 offspring each. Are populations A and B in the process of becoming different species?

**8.** A few plants of species $Q$ ($2n = 14$) suddenly double their chromosomes ($2n = 28$) and immediately become a new species, $R$. Why are $QR$ hybrids sterile?

**9.** One of the arguments creationists use to refute evolution is the presence of gaps in the fossil record. How can you explain the gaps from an evolutionary standpoint?

### GENETIC VARIATION

**10.** The following electrophoretic data are from a sample of one hundred field mice for their salivary amylase-1 genotypes. The two alleles are *F* and *S,* for fast and slow migration in an electric field: *FF,* forty-three, *FS,* fifty-four, and *SS,* three. Is selection acting? What would you look for in data to determine whether frequency-dependent selection, heterozygote advantage, or transient polymorphism is at work?

**11.** What mechanisms permit the maintenance of genetic variability in natural populations? Give examples where possible.

**12.** Discuss the "neutral gene hypothesis." What are its alternatives? What data are needed to distinguish among these views?

**13.** Koehn showed that differently functioning alleles of an esterase system in fish correlated with water temperature. What sorts of selection can you imagine that could affect the same type of alleles in mammals, which are homeothermic (warm-blooded) and hence maintain a relatively constant internal temperature?

**14.** P. Niemalä and J. Tuomi have suggested that the irregular leaf outlines in some plant species are a form of mimicry. What would the leaves be mimicking? What form of mimicry might this be?

**15.** From figure 21.10, what is $K$ (the average number of amino acid substitutions per site) between human beings and chickens? between dogs and human beings? Do all three possible comparisons support known evolutionary relationships?

**16.** How does the acceptance of the neutral mutation theory change our basic view of neo-Darwinism?

**17.** In a given population, the frequencies of *AA, Aa,* and *aa* genotypes are 0.36, 0.48, and 0.16, respectively. If the assigned fitnesses are $1.5 - p$, 1.0, and $1.5 - q$, what will the genotypic frequencies be after one generation of selection?

**18.** If the rate of amino acid substitution per site per year is $2 \times 10^{-9}$, and the average number of amino acid substitutions per site is 0.2, how long has it been since the two species diverged?

**19.** Scientists have examined one thousand amino acids in the proteins of human beings and chimps and have found a difference in twenty-three. Calculate the average number of amino acid substitutions per site.

---

*Answers to selected Exercises and Problems are on page A-24.

**20.** Scientists are now using DNA sequences to show phylogenetic relationships between or among species. In many cases, cDNA is made from isolated mRNA and then sequenced. Is the method a reasonable approach to show evolutionary relationships?

**21.** In which codon position should the greatest abundance of variation occur? Why?

**SOCIOBIOLOGY**

**22.** What are the differences among individual selection, group selection, and kin selection? How could each type of selection explain altruistic acts?

**23.** If the "calculus of the genes" suggests sacrificing oneself for two siblings, for how many first cousins should one sacrifice oneself?

**24.** In certain animal populations, infanticide is practiced by one or more males. Do you think this infanticide is random, or would you expect specific individuals to be eliminated?

## CRITICAL THINKING QUESTIONS

**1.** The peppered moth (*Biston betularia*) has two phenotypic forms, melanic (dominant) and peppered (recessive). The moths face predation by birds, and the predation is selective against different-colored tree trunks. In an industrialized area, one in which the tree trunks are dark like the melanic form (and thus "hide" the melanic forms from the birds), a sample of moths indicated that the frequency of the allele for peppering was 0.6; the next year, it was 0.5. What is the fitness of the peppered genotype under this circumstance?

**2.** V. C. Wynne-Edwards suggested that birds form flocks so that they can assess their population numbers. When they assess that their numbers are high, they decide not to breed for the good of the species, so that they do not exhaust their resources. Edwards called this process *group selection.* Why can't this mechanism work, given that it involves behavior that is for the good of the species?

*Suggested Readings for chapter 21 are on page B-20.*

# APPENDIX A

## Brief Answers to Selected Exercises, Problems, and Critical Thinking Questions

### Chapter 2 Mendel's Principles

**1.** Dwarf $F_2$ (1/4 of total $F_2$), when selfed, produce all dwarf progeny (*tt*). Tall $F_2$ (3/4 of total $F_2$), when selfed, fall into two categories: 1/3 (*TT,* 1/4 of total $F_2$) produce all tall, and 2/3 (*Tt,* 1/2 of total $F_2$) produce tall and dwarf progeny in a 3:1 ratio. (The 3:1 ratio is from 1/2 the $F_2$, so the tall component is 3/8 of the total $F_3$ [$3/4 \times 1/2$], and the dwarf is 1/8 of the total $F_3$ [$1/4 \times 1/2$].) Overall, the $F_3$ are 3/8 *TT* (tall), 2/8 *Tt* (tall), and 3/8 *tt* dwarf (see fig. 2.7).

**3.** Rule of segregation: adult diploid organisms possess two copies of each gene. Gametes get one copy. Fertilization restores the diploid number to the zygote. Rule of independent assortment: alleles of different genes segregate independently of each other.

**5.** Black is dominant, white is recessive, and both parents were heterozygous. The progeny are in an approximate 3:1 ratio. Since both parents had the same phenotype, the simplest cross is *Bb* × *Bb.*

**7.** The disease is recessive at the individual level but incompletely dominant at the enzymatic level. Check the glossary for definitions.

**9.** *Ll* × *Ll.* Let $L$ = long ears and $l$ = no ears. We see three phenotypes in an approximate 1:2:1 ratio. One of the phenotypes (short) is intermediate between long ears and no ears. Therefore, we have incomplete dominance.

**11.** Washed eye mutant, *We;* wild-type, $We^+$. (*W* is already the allelic designation for the wrinkled phenotype.)

**13.** All. Since the child was type A, it must have gotten the $I^A$ allele from its mother. The other allele in the child is either $I^A$ or *i.* A type A ($I^AI^A$ or $I^Ai$), type B ($I^BI^B$ or $I^Bi$), type O, or type AB man could have supplied either an $I^A$ or *i* allele.

**15.** Universal donor, type O (no red-cell antigens); universal recipient, type AB (no serum antibodies).

**17.** All AB; or 1/2 AB, 1/2 A; or 1/2 AB, 1/2 B; or 1/4 A:1/4 AB:1/4 B:1/4 O. Crosses can be $I^AI^A \times I^BI^B$, $I^AI^A \times I^Bi$, $I^Ai \times I^BI^B$, or $I^Ai \times I^Bi$.

$$\begin{array}{c} I^AI^A \times I^BI^B \\ \downarrow \\ \text{All } I^AI^B \\ \text{(AB)} \end{array} \qquad \begin{array}{c} I^Ai \times I^BI^B \\ \downarrow \\ 1/2\ I^AI^B{:}1/2\ I^Bi \\ \text{(AB)} \quad \text{(B)} \end{array}$$

$$\begin{array}{c} I^AI^A \times I^Bi \\ \downarrow \\ 1/2\ I^AI^B{:}1/2\ I^Ai \\ \text{(AB)} \quad \text{(A)} \end{array} \qquad \begin{array}{c} I^Ai \times I^Bi \\ \downarrow \\ 1/4\ I^AI^B{:}1/4\ I^Ai{:}1/4\ I^Bi{:}1/4\ ii \\ \text{(AB)} \quad \text{(A)} \quad \text{(B)} \quad \text{(O)} \end{array}$$

**19.** Steve and his fiancé could be related. Both the dean and Steve's father must be $I^Ai$ to produce O children, and each could have contributed *M* to produce M offspring. If the dean and Steve's father each contributed an *S* allele, the daughter would be *SS.* Note that if the daughter had B blood, she and Steve could not be related.

**21.** *RrTt* × self yields:

1/16 *RRTT* red, tall
2/16 *RRTt* red, medium
1/16 *RRtt* red, dwarf
2/16 *RrTT* pink, tall
4/16 *RrTt* pink, medium
2/16 *Rrtt* pink, dwarf
1/16 *rrTT* white, tall
2/16 *rrTt* white, medium
1/16 *rrtt* white, dwarf

**23.** Choice (b) is preferred because although each will give the correct genotype, generally, testcrossing has the greatest probability of exposing the recessive allele in a heterozygote. For example, an *Aa* genotype, when selfed, produces *aa* offspring one-fourth of the time; when testcrossed, *aa* offspring appear one-half of the time; and when crossed with the *Aa* type (backcross), *aa* offspring occur one-fourth of the time. Thus, with a limited number of offspring examined per cross, testcrossing most reliably exposes the recessive allele.

**25.** The $F_1$ are tetrahybrids (*Aa Bb Cc Dd*). If selfed, an $F_1$ would form $2^4 = 16$ different types of gametes; $2^4$ different phenotypes would appear in the $F_2$, which would be made up of $3^4 = 81$ different genotypes; $1/(16)^2 = 1/256$ of the $F_2$ would be of the *aa bb cc dd* genotype.

**27.** **a.** In the first cross, look at the yellow-to-green ratio, 120:43—almost exactly 3:1. Therefore, yellow is dominant and both parents must be heterozygous. Now look at the tall-to-short ratio, 122:41. Again, we see a 3:1 ratio, which indicates that tall is dominant and each parent is heterozygous. Thus, the cross is probably *YyTt* × *YyTt.*

**b.** In the second cross, there are no tall progeny. Therefore, either the short phenotypes are homozygous, or short is dominant and at least one parent is homozygous. In the absence of the first cross, we can't determine the mode of inheritance of height. We can, however, conclude that yellow is dominant (we got a 3:1 ratio) and that each parent is heterozygous. Based on the first cross, we can conclude that this cross is *Yytt* × *Yytt.*

**c.** In the third cross, we see 41 yellow:46 green, and 45 tall:42 short. Both of these ratios are 1:1, and all we can conclude is that these ratios result from matings between a heterozygote and a recessive homozygote. With only this cross, we can't determine dominance. However, we can if we use all three crosses. The cross is *yyTt* × *Yytt.*

**29.** From the $F_1$ progeny, we can see that long and tan must be dominant, and the $F_2$ result confirms this assumption. We see a 3:1 ratio for both tan:dark and long:short. The total number of flies is 80. An ideal 9:3:3:1 ratio would be 45:15:15:5. Our results are very close to this. Therefore, we conclude that tan and long are dominant, and that the $F_1$ flies were heterozygous.

**31.** First list all possible genotypes for colored plants:

*AACCRR*
*AACCRr*
*AACcRR*
*AACcRr*
*AaCCRR*
*AaCCRr*
*AaCcRR*
*AaCcRr*

The first genotype can be eliminated because all progeny should be colored, regardless of the tester strain. *AACCRr* can be eliminated because the progeny of the first cross would have all been colored. *AACcRR* can be eliminated because the progeny of the second cross would have all been colored. *AaCCRR* can also be eliminated because the progeny of the third cross would have all been colored.

We are now left with *AACcRr; AaCCRr; AaCcRR,* and *AaCcRr.* Try *AACcRr* × *aaccRR* (cross 1), which will give 1/2 colored (*A-C-R-*): 1/2 colorless (*A-ccR-*) progeny. This could be the genotype, so try it in the second cross: *AACcRr* × *aaCCrr.* This too will give 1/2 colored and 1/2 colorless offspring (*A-C-rr*). Since this does not fit the observed result, the unknown genotype is not *AACcRr.* Now try *AaCCRr* × *aaccRR* (cross 1). This fits the results. Try *AaCcRR* × *aaccRR.* This will give 3/4 colorless (*aaccRR, A-ccRR,* or *aaC-RR*), which does not fit the results; therefore, the genotype is not *AaCcRR.* Now try *AaCcRr* × *aaccRR.* This, too, will give 3/4 colorless (*aaccR-, A-ccR-, or aaC-R-*), which is not seen. Therefore, the genotype must be *AaCCRr.* Confirm this with the other two crosses:

*AaCCRr* × *aaCCrr*→1/4 colored : 3/4 colorless, which fits.

*AaCCRr* × *AAccrr*→1/2 colored : 1/2 colorless, which fits.

**33. a.** all normal

**b.** 9 normal: 7 dark

**c.** 1/2 ebony:1/2 normal; 1/2 black:1/2 normal

If we let $e$ = ebony, $e^+$ = wild-type, $b$ = black, and $b^+$ = wild-type, the first cross is:

$ee\,b^+b^+$ × $e^+e^+\,bb$
↓
all $e^+e\,b^+b$
↓ selfing

| | |
|---|---|
| 9/16 $e^+{-}b^+{-}$ | wild-type |
| 3/16 $ee\,b^+{-}$ | ebony |
| 3/16 $e^+{-}\,bb$ | black |
| 1/16 $ee\,bb$ | ebony, black |

Since it is difficult to distinguish black and ebony, 7/16 will be dark-bodied.

In c, the crosses are:

1. $e^+e\,b^+b \times ee\,b^+b^+$
2. $e^+e\,b^+b \times e^+e^+bb$

In each case we have a testcross situation for only one gene.

**35.** 9 wild-type:3 orange-1:3 orange-2:1 pink.

The first two crosses indicate that wild-type is dominant to both oranges, and the fourth indicates that orange-2 is dominant to pink. The fifth cross produces four phenotypes, indicating we are dealing with at least two genes. The presence of two genes is also suggested by orange-1 × orange-2. If these two traits were allelic, all the progeny should have been orange. The $F_1$ × pink produces progeny that resemble those from a double testcross. If *A-B-* = wild-type, *A-bb* = orange-1, *aaB-* = orange-2, *aabb* is probably pink. The crosses in question are then:

*AAbb* × *aaBB*
(orange-1) (orange-2)
↓
*AaBb*
(wild-type)
↓ self

| | |
|---|---|
| 9/16 *A-B-* | wild-type |
| 3/16 *A-bb* | orange-1 |
| 3/16 *aaB-* | orange-2 |
| 1/16 *aabb* | pink |

**37.** Two loci with epistasis. *AaBb* × self yields 9/16 *A-B-:6/16 A-bb* + *aaB-:1/16 aabb.* Verify by testcrossing the various classes.

**39.** See for example figure 2.25. Other ratios are 10:3:3; 10:6; 12:3:1; 12:4.

**41.** 9 nonworkers: 7 hard workers. The $F_1$ indicates nonworker is dominant; therefore, a worker must be a recessive homozygote. If one gene is involved, the cross of the $F_1$ female × worker male (*Ww* × *ww*) should produce 1 worker:1 nonworker, for this is a testcross. This result is not seen, so we must have more than one gene involved. Perhaps a worker can result from more than one gene. Let *A-B-* = nonworker, and *A-bb, aaB-,* or *aabb* = workers. The original worker is *aabb,* and the cross is:

*aabb* × *AABB*
↙
*AaBb* × *aabb*
(nonworker) (worker)

| 1/4 *A-B-:* | 1/4 *A-bb:* | 1/4 *aaB-:* | 1/4 *aabb* |
|---|---|---|---|
| nonworker | worker | worker | worker |

If *AaBb* × *AaBb*
↓
9/16 *A-B-* nonworkers

3/16 *aaB-*
3/16 *A-bb*
1/16 *aabb*
} workers

**43.** $? \xrightarrow{2} \text{indole} \xrightarrow{1} \text{tryptophan or kynurenine} \xrightarrow{3} \text{3-hydroxyanthranilic acid} \xrightarrow{4} \text{niacin}$

Accumulation: 1, indole; 2, ?; 3, tryptophan or kynurenine; 4, 3-hydroxyanthranilic acid. Without serine, the pathway would be blocked before the point of tryptophan production, after indole.

**45. a.** Maple sugar urine disease is recessive. If two individuals with the same phenotype produce offspring, some of whom have a different phenotype, both parents must be heterozygous. Since they are heterozygous, their phenotype must be dominant.

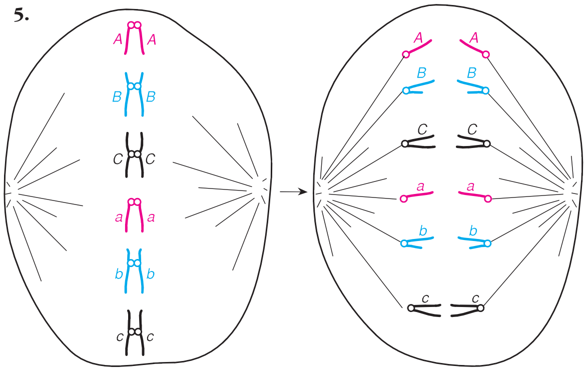
**b.** 3/4. Let *M* = normal and *m* = maple sugar urine alleles. The cross is *Mm* × *Mm.* At each conception, the probability is three-fourths that each child is of the *M-* genotype.

*Critical Thinking Question:*

1. The 15:1 ratio indicates that all genotypes except the recessive homozygote (*aabb*) produce a triangular capsule. The rounded capsule results from the recessive homozygote. One way to look at this is that the rounded form is a "default" form when neither locus has functional—dominant—alleles. However, a dominant allele at either of two loci is adequate to form the triangular seed capsule. The loci can be considered redundant in the pathway of seed capsule shape since a functional allele at one, the other, or both will provide a dominant (triangular) phenotype. At this point in time, it is impossible to know precisely what the enzymatic function of each dominant allele is.

### Chapter 3 Mitosis and Meiosis

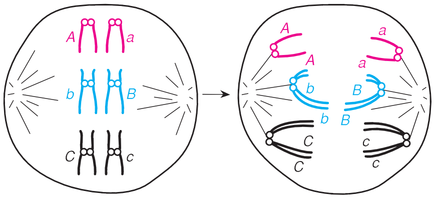
1. See table 3.1 for a summary answer.
3. When a eukaryotic chromosome replicates during the S phase of the cell cycle, one chromosome becomes two chromatids, attached near the centromere. These are sister chromatids. Chromatids of different chromosomes are nonsister chromatids. Homologous chromosomes are members of a pair of essentially identical chromosomes. In diploid organisms, one member from each pair comes from each parent. Nonhomologous chromosomes do not share this relationship.
5.



Metaphase, mitosis

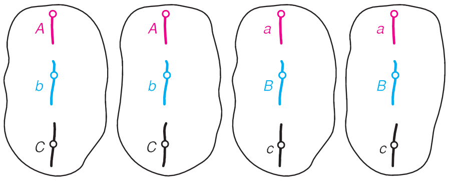
Anaphase, mitosis

7. S phase
9. $2^3 = 8$ different gametes can arise. A crossover between the *A* locus and its centromere does not alter gametic combinations.



Metaphase I, meiosis

One possible anaphase I, meiosis



Telophase II, meiosis

11. The intent of this problem is to make you think about the essential steps of meiosis, primarily the necessity to separate members of homologous pairs of chromosomes. Presumably, any method you devise will force you through that process.
13. Meiosis apportions homologous chromosomes the same way Mendel's rules apportion alleles. Each gamete gets one member of a homologous pair of chromosomes. Segregation predicts the same about alleles. The separation of homologues of one chromosome pair at meiosis is independent of the separation of other homologous pairs. Independent assortment makes the same prediction about alleles.
15. A gamete from wheat will have twenty-one chromosomes, and a gamete from rye will have seven chromosomes. Even if the seven rye chromosomes could pair with seven wheat chromosomes, a highly unlikely possibility, the remaining fourteen wheat chromosomes could not pair and would segregate randomly during meiosis. Almost every gamete would get an incomplete set; if fertilization did occur, the zygotes would have extra chromosomes (trisomic) or would be missing some chromosomes (monosomic or nullosomic).
17. 64. The number of combinations is $2^n$ where $n$ = the number of different chromosomes in the set.
19.

| | **DNA (Number of Chromatids)** | **Ploidy** |
|---|---|---|
| Spermatogonium or Oogonium | 2 | $2n$ |
| Primary Spermatocyte or Primary Oocyte | 4 | $2n$ |
| Secondary Spermatocyte or Secondary Oocyte | 2 | $n$ |
| Spermatid or Ovum | 1 | $n$ |
| Sperm | 1 | $n$ |

21. **a.** 50 **b.** 50. The primary oocyte is diploid and will undergo meiosis, but only one functional ovum results from each primary oocyte. The secondary oocyte will divide to produce an ovum and a polar body.
23. Any possible genotype, from *AAABBB* through *aaabbb* can occur in the endosperm. If at a given locus the endosperm is homozygous, so is the embryo. If the locus is heterozygous (e.g., *AAa* or *Aaa*), so is the embryo. Thus, an *AAabbb* endosperm is associated with an *Aabb* embryo.
25. A greater maternal influence in *Drosophila* and corn than in *Neurospora* means the sexes (mating types) do not show the disparity in size between male and female cells as in *Drosophila* and corn.
27. **b.** Homologous chromosomes will pair during meiosis. Each gamete gets one of each chromosome, A, B, C, D, and E. Fertilization fuses two cells with the chromosome complement given. Since root cells are somatic tissue, these cells will be diploid.
29. **a.** $2^{50}$ or about $1.1 \times 10^{15}$ **b.** $2^2$. The number of gametes produced is $2^n$, where $n$ = number of independently behaving entities. If the genes are completely independent, we expect $2^{50}$, and if they are completely linked, we expect $2^2$. In reality, the number falls between these extremes.

*Critical Thinking Question:*

1. Both meiosis and mitosis are processes that initiate under certain circumstances of cell cycle and place. Neither is actually dependent on the chromosomal content of the cell. Thus, meiosis could begin in a haploid cell, but it would not be a successful process because there is no homologue for any chromosome to pair with. Mitosis would, however, be successful because there is no pairing (synapsis) required for successful completion of the process.

## Chapter 4 Probability and Statistics

1. **a.** $(5!/3!2!)(1/2)^3(1/2)^2 = 0.3125$
   **b.** $(1/2)^5 = 0.03125$ (SDSDS, in which S = son, D = daughter)
   **c.** $2(1/2)^5 = 0.0625$ (SDSDS + DSDSD)
   **d.** $(1/2)^5 = 0.03125$
   **e.** $2(1/2)^5 = 0.0625$ (all sons + all daughters)
   **f.** 4 daughters, 1 son + 5 daughters: $(5!/4!1!)(1/2)^4(1/2) + (1/2)^5 = 0.1875$
   **g.** $(1/2)^2 = 0.25$ (DXXXS, in which X is either a daughter or a son, with $p = 1$; $P = [1/2][1][1][1][1/2]$)
3. Remember that albinos have blue eyes. Therefore, 7/16 of the offspring will have blue eyes. If we let $B$ = brown, $b$ = blue, $C$ = normal color expression, and $c$ = albinism, the following genotypes are blue-eyed: *C-bb* and *cc–*.
   **a.** $(1/4)^5 = 0.0009765$
   **b.** $(1/8)^5 = 0.0000305$
   **c.** $(5!/4!1!)(7/32)^4(9/32) = 0.00322$
   **d.** $(4!/2!2!)(1/8)^2(1/8)^2 = 0.0014648$
5. **a.** $2(1/2)^4 = 0.125$
   **b.** $(1/2)^4(1/2)^4 = 0.0039063$ (Probability that sperm and egg creating the zygote each had only paternal centromeres.)
7. One-half.
9. **a.** 81/256 **b.** 108/256 **c.** 9/256. In (a), since all children have the same phenotype, each child will have the same probability of having no molars. Therefore, $(3/4)^4 = 81/256$. In (b),

$$P = \frac{4!}{3!1!}(3/4)^3(1/4) = 108/256.$$

   When order is given, we multiply the chance of each event, $1/4 \times 1/4 \times 3/4 \times 3/4 = 9/256$.
11. One-eighth. B must be heterozygous (*Gg*), as must A's father. We assume A's mother is *GG*, since there is no mention of the disease in her family. Therefore, A has a one-half chance of getting *g* from his father. If two heterozygotes mate, the chance of a recessive child is one-fourth, so $P = 1/4 \times 1/2$.
13. 1/512. The $F_1$ progeny are *Aa Bb Cc Dd Ee*. The chance of getting any individual with a particular homozygous genotype is $(1/4)^5$. Since we are looking for two different possibilities, we have $2(1/4)^5 = 2/1024 = 1/512$.
15. 0.049. Since the order is not specified, we use the multinomial formula. We have six mice, so $n = 6$. If $p$ = chance of agouti, $p = 3/4 \times 1/2 = 3/8$; $q$ (black coat color) $= 1/4 \times 1/2 = 1/8$, and $r$ (albino) $= 1/2$. The equation becomes:

$$p = \frac{6!}{2!2!2!}(3/8)^2(1/8)^2(1/2)^2 =$$

$$\frac{6 \times 5 \times 4 \times 3 \times 2 \times 1 \times 9 \times 1 \times 1}{2 \times 2 \times 2 \times 64 \times 64 \times 4} = 0.049$$

17. Hypothesis: *RrYy* × *RrYy* produces *R-Y-*:*R-yy*:*rrY-*:*rryy* in a 9:3:3:1 ratio. The critical chi-square, three degrees of freedom at probability of 0.05, = 7.815.

| | ***R-Y-*** | ***R-yy*** | ***rrY-*** | ***rryy*** | **Sum** |
|---|---|---|---|---|---|
| *Observed* | 315 | 108 | 101 | 32 | 556 |
| *Expected* | 9/16 | 3/16 | 3/16 | 1/16 | |
| | 312.75 | 104.25 | 104.25 | 34.75 | 556 |
| $O - E$ | 2.25 | 3.75 | −3.25 | −2.75 | |
| $(O - E)^2$ | 5.06 | 14.06 | 10.56 | 7.56 | |
| $(O - E)^2/E$ | 0.016 | 0.135 | 0.101 | 0.218 | $0.470 = \chi^2$ |

   Since this chi-square, 0.470, is less than the critical chi-square, we fail to reject our hypothesis of two-locus genetic control with dominant alleles at each locus.
19. We reject the 3:1 ratio as an appropriate null hypothesis. If we calculate the chi-square using 3:1 as the expected ratio, we expect 72 and 24 (3/4 and 1/4 of 96, respectively):

| | ***O*** | ***E*** | ***O–E*** | $(O\text{–}E)^2$ | $(O\text{–}E)^2/E$ |
|---|---|---|---|---|---|
| Curly-winged flies | 61 | 72 | −11 | 121 | 121/72 = 1.681 |
| Straight-winged flies | 35 | 24 | 11 | 121 | 121/24 = 5.042 |

   Chi-square = 6.723 (1.681 + 5.042). With one degree of freedom, $p > 0.05$ (critical chi-square = 3.841).

*Critical Thinking Question:*

1. You should change your choice because the box you chose originally has a 1/3 chance of containing the prize, whereas the remaining box has a 2/3 chance of containing the prize. The 1/3 chance of your choice is set by the fact that there were three equally likely choices at the beginning. When your friend eliminated an empty box, she left two choices: your original box and the third box. Since the probability of your original choice has not changed, the probability of the remaining box must be 2/3 to give a combined probability of 1.0 that a box contains a prize.

## Chapter 5 Sex Determination, Sex Linkage, and Pedigree Analysis

1. The differences are in terminology only, not in shape or size of the chromosomes. In species in which females have a homomorphic pair of sex chromosomes, the members of the pair are called X chromosomes. In species in which males have a homomorphic sex chromosome pair, the members of the pair are called Z chromosomes.
3. 3/8 males, 3/8 females, 2/8 intersexes (*dsx dsx* homozygotes). The 1/4 *dsx dsx* flies will be intersexes, whereas half of the 3/4 will be normal males and half will be normal females.
5. The protein is probably a dimer, which, in the heterozygote, can be of fast-fast, fast-slow, or slow-slow subunit combinations. A female heterozygous for a sex-linked gene controlling a dimeric enzyme should show the pattern of lane 3 in whole blood (mixture of slow-slow and fast-fast dimers) and lanes 1 or 2 in individual cells.

**7.** **a.** 0; human female, male fly
- **b.** 1; human female, female fly
- **c.** 0; human male, male fly
- **d.** 1; human male, female fly
- **e.** 2; human female, female fly
- **f.** 4; human female, female fly
- **g.** 1/0 mosaic; human male-female mosaic, male-female fly mosaic

**9.**

| | **Cross** | **Reciprocal** |
|---|---|---|
| *$P_1$* female | $X^+X^+$ | $X^{lz}X^{lz}$ |
| male | $X^{lz}Y$ | $X^+Y$ |
| *$F_1$* female | $X^+X^{lz}$ | $X^+X^{lz}$ |
| male | $X^+Y$ | $X^{lz}Y$ |
| *$F_2$* females | $X^+X^+$, $X^+X^{lz}$ | $X^+X^{lz}$, $X^{lz}X^{lz}$ |
| males | $X^+Y$, $X^{lz}Y$ | $X^+Y$, $X^{lz}Y$ |

**11.** Exemptions should be made minimally for hemophilia in brother, sister's son, mother's brother, mother's sister's son, mother's father, and others more distantly related.

**13.** $P_1$ $fyfy\ X^+X^+$ (female) × $fy^+fy^+\ X^{ct}Y$ (male)

$F_1$ $fy^+fy\ X^+X^{ct}$ (female) × $fy^+fy\ X^+Y$ (male)

| | **Male** | | | |
|---|---|---|---|---|
| | ***$fy^+\ X^+$*** | ***$fy\ X^+$*** | ***$fy^+\ Y$*** | ***$fy\ Y$*** |
| Female | | | | |
| $fy^+\ X^+$ | $fy^+fy^+\ X^+X^+$ | $fy^+fy\ X^+X^+$ | $fy^+fy^+\ X^+Y$ | $fy^+fyX^+Y$ |
| $fy^+\ X^{ct}$ | $fy^+fy^+\ X^+X^{ct}$ | $fy^+fy\ X^+X^{ct}$ | $fy^+fy^+\ X^{ct}Y$ | $fy^+fy\ X^{ct}Y$ |
| $fy\ X^+$ | $fy^+fy\ X^+X^+$ | $fyfy\ X^+X^+$ | $fy^+fy\ X^+Y$ | $fyfy\ X^+Y$ |
| $fy\ X^{ct}$ | $fy^+fy\ X^+X^{ct}$ | $fyfy\ X^+X^{ct}$ | $fy^+fy\ X^{ct}Y$ | $fyfy\ X^{ct}Y$ |

$F_2$: females, 3/4 wild-type, 1/4 fuzzy; males, 3/8 wild-type, 3/8 cut, 1/8 fuzzy, and 1/8 cut and fuzzy.

**15.** **a.** X linked
- **b.** gray
- **c.** 1/2 gray:1/2 yellow in both sexes

In both crosses, we see a difference in the phenotypes of the sexes, suggesting sex linkage. The $F_1$ offspring from the first cross indicate that gray is dominant to yellow. The $F_1$ females from this cross must be heterozygous, and the two phenotypes in the $F_2$ males result from each of the X chromosomes in the $F_1$ female being hemizygous in the $F_2$ males. The first cross is therefore (calling gray the wild-type):

$$\begin{array}{llll} X^+X^+ & \times & X^yY & \\ & \downarrow & & \\ X^+X^y & \times & X^+Y & \\ \text{gray} & & \text{gray} & \\ & \downarrow & & \\ X^+X^+ & X^+X^y & X^+Y & X^yY \\ \text{gray} & \text{gray} & \text{gray} & \text{yellow} \end{array}$$

Now diagram the second cross:

$$\begin{array}{llll} X^yX^y & \times & X^+Y & \\ & \downarrow & & \\ X^+X^y & \times & X^yY & \\ \text{gray} & & \text{yellow} & \\ & \downarrow & & \\ X^+X^y & X^yX^y & X^+Y & X^yY \\ \text{gray} & \text{yellow} & \text{gray} & \text{yellow} \end{array}$$

**17.** Yes. Begin by determining genotypes of the two individuals. The woman must be heterozygous $X^CX^c$. A man with normal vision must be $X^CY$, and all his daughters must receive his X chromosome and should be normal, either $X^CX^C$ or $X^CX^c$. Since color blindness is recessive, the daughter must have two $X^c$ chromosomes. A *very* rare possibility is that the man is the father and nondisjunction occurred in both parents: at meiosis I in the male and at meiosis II in the female (see chapter 8).

**19.** **a.** $F_1$: wild-type females, white-eyed males.
- **b.** $F_2$: 3 wild-type:3 white-eyed:1 ebony:1 ebony, white-eyed in both sexes.
- **c.** Reciprocal $F_2$: 3 wild-type females:1 ebony female:3 wild-type males:3 white-eyed males:1 ebony male:1 ebony, white-eyed male. Let $X^+$ = red, $X^w$ = white, $e^+$ = wild-type, $e$ = ebony.

$$\begin{array}{llll} X^wX^w\ e^+e^+ & \times & X^wY\ ee & P_1 \\ & \downarrow & & \\ X^+X^w\ e^+e & \times & X^wY\ e^+e & F_1 \\ \text{(wild-type)} & & \text{(white-eyed)} & \end{array}$$

Use probability for the $F_2$ generation rather than the Punnett square:

$$\begin{array}{rl} (1/2)X^+ \times (1/2)X^w \times & (3/4)e^+\text{-} = 3/16 \text{ wild-type females} \\ \times & (1/4)ee\ \ = 1/16 \text{ ebony females} \\ \times (1/2)Y \times & (3/4)e^+\text{-} = 3/16 \text{ wild-type males} \\ \times & (1/4)ee\ \ = 1/16 \text{ ebony males} \\ (1/2)X^w \times (1/2)X^w \times & (3/4)e^+\text{-} = 3/16 \text{ white-eyed females} \\ \times & (1/4)ee\ \ = 1/16 \text{ white-eyed, ebony females} \\ \times (1/2)Y \times & (3/4)e^+\text{-} = 3/16 \text{ white-eyed males} \\ \times & (1/4)ee\ \ = 1/16 \text{ white-eyed, ebony males} \end{array}$$

For the reciprocal cross,

$$\begin{array}{lll} X^+X^+\ ee & \times & X^wY\ e^+e^+ \\ & \downarrow & \\ X^+X^w\ e^+e & \times & X^+Y\ e^+e \\ \text{(wild-type)} & & \text{(wild-type)} \end{array}$$

All $F_2$ females will get $X^+$; $e^+$:$e$ will be 3:1. The males will be as in the males shown.

**21.** The female is heterozygous for an X-linked color gene (one of the X chromosomes in the cells of female cats is inactivated, leading to the black and yellow spots). We immediately deduce X-linkage because of the different phenotypes in the sexes. Finding two types of males indicates that the female was heterozygous. The patches of yellow and black come from X-inactivation. The cross is:

$$\begin{array}{lllll} & X^bX^o & \times & X^bY & \\ & & \downarrow & & \\ X^bX^b & X^bX^o & & X^bY & X^oY \\ \text{black} & \text{black} & & \text{black} & \text{yellow} \\ & \text{and yellow} & & & \end{array}$$

**23.** Penetrance is the proportion of individuals of a particular genotype that shows the appropriate phenotype; expressivity is the degree to which a trait is expressed.

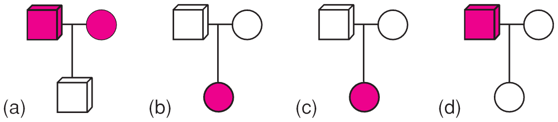
**25. a.** The phenotype is the propensity to have twin offspring. It could be caused by a recessive or dominant, sex-linked or autosomal allele.

**b.** Autosomal dominant or possibly autosomal recessive inheritance.

**c.** Autosomal, or sex-linked, recessive inheritance.

**d.** Autosomal recessive inheritance.

**27.** Assuming 100% penetrance:



*Critical Thinking Question:*

**1.** The immediate effect of a null allele is to make heterozygous genotypes appear to be homozygotes. That is, in the simplest system, we expect one band in a homozygote and two bands in a heterozygote. If we see only one band, we assume that the individual is homozygous for that allele, when in fact that individual might be heterozygous for the allele that produces the particular band and the allele that produces no band (null allele). The null allele can be verified by both the absence of any bands in the null homozygote and the results of breeding experiments when the null allele is suspected.

### Chapter 6 Linkage and Mapping in Eukaryotes

**1. a.**

| | | |
|---|---|---|
| $P_1$ groucho | × | rough |
| $grogro\ ro^+ro^+$ | | $gro^+gro^+\ roro$ |
| $F_1$ female $gro^+gro\ ro^+ro$ | × | male $grogro\ roro$ |
| $F_2$ $grogro\ ro^+ro$ | | 518 |
| $gro^+gro\ roro$ | | 471 |
| $grogro\ roro$ | | 6 |
| $gro^+gro\ ro^+ro$ | | 5 |

(6+5)/1,000 = 0.011 = 1.1% recombination
= 1.1 map units apart

**b.** Given the map units, $F_1$ gametes are produced on the average by females as follows: $gro\ ro^+$, 49.45% = 0.4945 (98.9%/2); $gro^+\ ro$, 49.45% = 0.4945; $gro\ ro$, 0.55% (1.1%/2) = 0.0055; and $gro^+\ ro^+$, 0.55% = 0.0055. Males, lacking crossing over, produce only two gamete types: $gro\ ro^+$ and $gro^+\ ro$, each 50% = 0.50. Summing from the Punnett square following, the phenotypes of the offspring would be as follows: wild-type, 50%; groucho, rough, 0%; groucho, 25%; and rough, 25%.

| | **Male** | |
|---|---|---|
| **Female** | **$gro\ ro^+$ (0.5)** | **$gro^+\ ro$ (0.5)** |
| $gro\ ro^+$ | groucho | wild-type |
| (0.4945) | 0.24725 | 0.24725 |
| $gro^+\ ro$ | wild-type | rough |
| (0.4945) | 0.24725 | 0.24725 |
| $gro\ ro$ | groucho | rough |
| (0.0055) | 0.00275 | 0.00275 |
| $gro^+\ ro^+$ | wild-type | wild-type |
| (0.0055) | 0.00275 | 0.00275 |

**3.** A dihybrid female is testcrossed (with a hemizygous male having both recessive alleles). Each recombinant class will make up about 5% of the offspring. Each parental class will make up about 45% of the offspring. Phenotypic classes will be equally distributed between the two sexes. The same results will be found for an autosomal locus if the dihybrids are females (no crossing over in males). A reciprocal cross cannot be done for X-linked genes because males cannot be dihybrid. Males dihybrid for an autosomal gene produce only two classes of offspring when testcrossed—parentals.

**5. a.** The hotfoot locus is in the middle (compare, for example, hotfoot, a double crossover, with the wild-type, a parental); there are 16.0 map units from hotfoot to either end locus: 74 + 66 + 11 + 9 recombinants between hotfoot and waved and 79 + 61 + 11 + 9 recombinants between hotfoot and obese.

**b.** The trihybrid parent was $o\ h\ wa/o^+h^+\ wa^+$.

**c.** The coefficient of coincidence is 20/25.6 (20/[0.16 × 0.16 × 1,000]); interference is 1 − (20/25.6) = 0.22, or 22%.

**7. a.** Work backward from the 0.61% double recombinants (0.100 × 0.061 × 100). Thus, there would be 6 of 1,000 double recombinants. In the *an-sple* region, we need the total of single + double recombinants = 100 of 1,000 (10 map units). Thus, 100 − 6 = 94; divided by 2 (two phenotypes) is 47 each. For the *spleat* region, the total of single and double recombinants = 61 (6.1 map units). Thus, 61 − 6 = 55; divided by 2 is 27 and 28. The parentals make up the remainder for a total of 1,000.

**b.** With a coefficient of coincidence of 0.60, only 0.366% (0.61 × 0.60) of the expected double recombinants will occur, that is, 4 instead of 6. Thus:

| | **Coefficient of Coincidence** | |
|---|---|---|
| | **1.0** | **0.6** |
| ancon, spiny, arctus oculus | 422 | 421 |
| wild-type | 423 | 422 |
| ancon, spiny | 27 | 28 |
| arctus oculus | 28 | 29 |
| ancon | 47 | 48 |
| spiny, arctus oculus | 47 | 48 |
| ancon, arctus oculus | 3 | 2 |
| spiny | 3 | 2 |
| Total | 1,000 | 1,000 |

**9. a.** linked; **b.** *trans;* **c.** 28.7%. The cross is a testcross. If the genes were not linked, we would expect a 1:1:1:1 ratio of offspring; we don't see that. The alleles that are linked will appear as the majority classes, which are Trembling, long-haired and normal, Rex. Therefore, Trembling and Rex are in the *trans* position. If we let $T$ = Trembling, and $R$ = Rex, the cross is

$$\frac{Tr}{tR} \times \frac{tr}{tr}$$

Recombinants are Trembling, Rex and normal, long-haired;

$$\frac{42 + 44}{300} \times 100 = 28.7\%$$

**11. a.** $\frac{k\ e^+\ cd}{k^+\ e\ cd^+}$ **b.** $\underline{k \quad 6.9 \quad e \quad 5.1 \quad cd}$

The initial cross is $\frac{k\ e^+cd}{k\ e^+cd} \times \frac{k^+e\ cd^+}{k^+e\ cd^+}$

Producing a trihybrid $F_1$ female: $\frac{k\ e^+\ cd}{k^+\ e\ cd^+}$, in any order.

The last two classes (3 + 4 offspring in the $F_2$) are double crossovers and allow us to determine order by comparison with the parentals (880 + 887). If the order is *k cd e,* a double crossover in the $F_1$ females yields $k\ cd^+\ e^+$ and $k^+\ cd\ e$. If the order is *cd k e,* a double crossover yields $cd\ k^+\ e^+$ and $cd^+\ k\ e$. Therefore, the order must be *k e cd,* which gives the correct double recombinants in the $F_2$ generation. After reconstructing the trihybrid

$$\frac{k\ e^+\ cd}{k^+\ e\ cd^+}$$

and scoring each of the offspring for crossovers in the *k-e* and *e-cd* regions: map units, *k-e* = ([64 + 67 + 4 + 3]/2000) × 100 = 6.9 and map units, *e-cd* = ([49 + 46 + 4 + 3]/2000) × 100 = 5.1.

**13.** 0.0125. This problem requires the manipulation of equations. We know that interference = 1 − coefficient of coincidence, so coefficient of coincidence = 1 − interference = 1 − (−1.5) = 2.5. Since coefficient of coincidence =

$$\frac{\text{observed double crossovers}}{\text{expected double crossovers}}$$

observed double crossovers = (coefficient of coincidence) × (expected double crossovers). The expected double crossover frequency is (2.5)(0.005) = 0.0125.

**15.** PD, 1, 2, 4, 6, 8-10; NPD, 3; TT, 5, 7. Map units = ([NPD + {1/2}TT]/Total) × 100 = ([1 + {1/2}2]/10) × 100 = 20. The loci are 20 map units apart.

**17.** FDS, 3-5, 8, 10; SDS, 1, 2, 6, 7, 9. The distance between the *arg* locus and its centromere is (1/2)%SDS = (1/2)50% = 25 map units.

**19.** For example, the only variant of the type 2 pattern of table 6.7 is $a^+b$, $a^+b$, $a^+b$, $a^+b$, $ab^+$, $ab^+$, $ab^+$, $ab^+$. Other patterns that are variants of the remaining five categories are derived by inverting the eight spores of a pattern (bottom to top) or by switching spores 1 and 2 with spores 3 and 4 or spores 5 and 6 with spores 7 and 8.

**21.** $a \times a^+ \rightarrow a/a^+$, which undergoes meiosis. Twelve map units means that the SDS pattern makes up 24% of the asci.

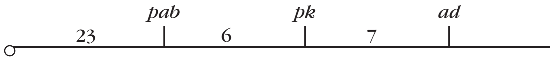
**23.** For example: 1, 408; 2, 42; 3, 250; 4, 250; 5, 30; 6, 5; 7, 15. To make the numbers work out, classes 4-7 must equal 300, as must classes 3 + 5−7, in order to make (1/2)%SDS = 150. Thus, if classes 3 and 4 are 250 each, 50 must be spread out among classes 5-7. NPD + (1/2)TT should equal 300 to confirm the arrangement (*a-b* distances). These numbers will give an *a-b* distance of 30.

**25.** **a.** yes; PD > > NPD (actually, no NPD). **b.** *a:* 2.5 map units; *b:* 7.5 map units. Classify each ascus—I: PDT, FDS for both; II: TT, FDS for *a,* SDS for *b;* III: TT, FDS for *a,* SDS for *b;* IV: TT, FDS for *a,* SDS for *b;* V: TT, FDS for *a*, SDS for *b;* VI: PDT, SDS for *a* and *b;* VII, PDT, SDS for *a* and *b;* VIII: PDT, SDS for *a* and *b.* We see no NPDs, so genes are linked. For gene to centromere distances, use the formula

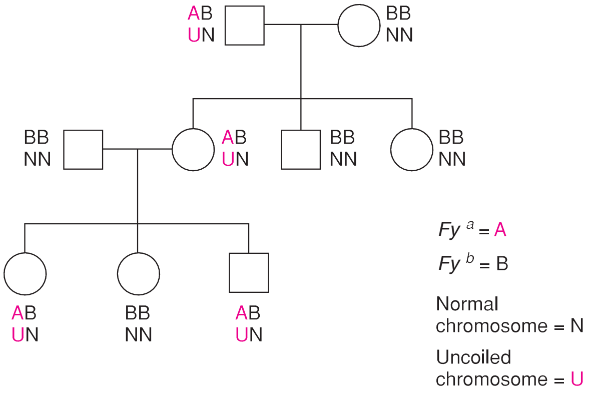
$$\frac{1/2\ (\text{number of SDS asci})}{100} \times 100:$$

*a* to centromere = (1/2)(3 + 1 + 1)% = 2.5 map units; *b* to centromere = (1/2)(2 + 3 + 2 + 3 + 3 + 1 + 1)% = 7.5 map units.

**27.** 12.5 map units. The only genotype that grows on minimal medium is $arg^+ade^+$. If the two genes were unlinked, 1/4 of the progeny should have this genotype; this is not seen. The genes must be linked; wild-type results from recombination between these two genes. The reciprocal class, $arg^-ade^-$, which has not been selected for, should be equally frequent, so: map units = (2 × 25)/400 × 100 = 12.5.



**29.** Construct a pedigree of the Duffy alleles ($Fy^a$, $Fy^b$). Arbitrarily assign one allele to a normal chromosome 1 and the other allele to the coiled chromosome. Then, accompanying the pedigree, the alleles and their morphologically proper chromosomes would be associated.



**31.** Twenty map units apart (two of ten recombinant sons according to the “grandfather method”).

**33.** Thirty-three map units. The woman is heterozygous in *trans* configuration for color blindness and hemophilia: $bc^+/b^+c$. Recombination between these two markers yields $b^+c^+$ and *bc.* We can only detect recombinants in sons, so

$$\frac{\#\ \text{normal sons}\ +\ \#\ \text{double mutant sons}}{\text{total sons}} \times 100$$

$$= \text{map distance} = 2/6 = 0.33$$

**35.** Enzyme A on 11; B on 15; C on 18; D on 3; E on 7. Enzyme A is present in clones X and Y, and chromosome 11 is common to these two clones. Enzyme B is present only in X, and 15 is the only chromosome unique to X. Similar logic allows the assignment of the other genes.

*Critical Thinking Question:*

**1.** Three-point crosses capture (allow us to see) double crossovers that have taken place in the two regions defined by the three loci. However, any double crossovers that occur within one region or any crossovers involving more than two events will not be indicated correctly by random-strand analysis.

## Chapter 7 Linkage and Mapping in Prokaryotes and Bacterial Viruses

**1.** The prokaryotic chromosome is a double-stranded DNA circle that is small compared with most eukaryotic chromosomes. Viral chromosomes can be DNA or RNA. Viruses are obligate intracellular parasites. Whether they are alive depends on the definition of the term *alive.*

**3.** A colony is a visible mass of cells derived usually from a single progenitor. A plaque is the equivalent growth of phages on a bacterial lawn, producing a cleared area lacking intact bacteria.

**5.** The bacterium could have survived and produced a colony if it was on a λ-free area, it became lysogenic (and thus resistant to further phage attack), or it was genetically resistant to phage λ.

**7.** 1, $his^-$ $arg^-$; 2, $leu^-$; 3, $lys^-$; 4, $his^-$ $met^-$ or $his^+$ $met^-$; 5, $arg^-$.

**9.** Where phages cannot grow: *E. coli* $ton^r$, phage $h^+$. Where phage can grow: *E. coli* $ton^s$, phage $h^+$ or *h,* or *E. coli* $ton^r$, phage *h.*

**11.** Far. If the selective locus is near, it passes into the $F^-$ cell very early during conjugation. Consequently, there is a great reduction in the recovery of loci distal to the selective marker because both the Hfr and $F^-$ members of a conjugation event can be killed by the selective agent (e.g., an antibiotic such as streptomycin).

**13.** See figures 7.8, 7.9, 7.15, 7.17, 7.18, and 7.26.

**15.** 9 min 1 min 8 min 7 min

Origin ——— *az* ——— *ton* ——— *lac* ——— *galB*

**17.** For example, use an Hfr that is wild-type but $str^s$ with the F factor integrated at minute 20. Use an $F^-$ strain that is $pyrD^-$, $purB^-$, $man^-$, $uvrC^-$, $his^-$, and $str^r$. Interrupt mating at one-minute intervals; plate cells on complete medium with streptomycin to kill Hfr cells and grow recombinant and nonrecombinant $F^-$ cells. The next day, after colonies have grown up, replica-plate onto selective media. The following data would be generated:

| | | Colony Growth on Media Selective for | | | | |
|---|---|---|---|---|---|---|
| | | $pyrD^+$ | $purB^+$ | $man^+$ | $uvrC^+$ | $his^+$ |
| Minute | 0 | − | − | − | − | − |
| | 1 | + | − | − | − | − |
| | 5 | + | + | − | − | − |
| | 16 | + | + | + | − | − |
| | 22 | + | + | + | + | − |
| | 24 | + | + | + | + | + |

**19.** The order is *a c b,* and *c* is close to *a.* Genes *c* and *a* are cotransformed 76% of the time, suggesting that these two genes are very close and *b* is far away (*a-b* cotransduction is 10%, or 0.1). Two orders are possible: *a-c*————*b* and *c-a*————*b.* If the first order is correct, $a^+\ b^+\ c^-$ results from a double crossover; this class should be the least frequent. If the second order is correct, a single exchange between *a* and *c* would yield $a^+\ b^+\ c^-$, but this frequency should be similar to $a^+\ b^-\ c^-$, and it is not.

**21.** *thr leu pro his.* We see that cells that are $thr^+$ are the most frequent. The chance of interruption in the conjugation increases with the length of time for the mating. Therefore, genes farther from the origin of transfer appear less frequently. We can order the genes based merely upon the frequency of genotypes seen. The order must be *thr leu pro his.* Since we see no $his^+$, and since we stopped the mating at 25 minutes, *his* must be after minute 25 on the map of this Hfr strain.

**23.** *a* and *c* are close; *b* is farther away; *c* is probably in the middle. The numbers of the first three transformant classes indicate that each gene, by itself, is readily transformed. We notice that classes with *b* and any other gene are quite rare, a situation that indicates *b* is far from *a* and *c.* We notice that *a* and *c* are cotransformed about 13% of the time, *b* and *c* about 3% of the time, and *a* and *b* about 2% of the time.

**25. a.** $lys^+\ his^+\ val^+$

$lys^+\ his^+\ \mathrm{val}^-$

$lys^+\ his^-\ val^+$

$lys^+\ his^-\ val^-$

Since there is no lysine in the medium, $lys^+$ must be present to allow growth.

**b.** $lys^+\ val^+\ his^+$

$lys^+\ val^+\ his^-$

Both $lys^+$ and $val^+$ must be present to allow growth.

**c.** $lys^+\ val^+\ his^+$

$lys^+\ val^-\ his^+$

Both $lys^+$ and $his^+$ must be present to allow growth.

**d.** $lys^+\ val^+\ his^-$

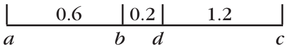
$lys^+\ val^-\ his^+$

We see no $lys^+\ val^+\ his^+$ cells.

**e.** $lys^+$ and $val^+$ are close together; they are cotransformed 75% of the time. Order could be *lys val his* or *val lys his.*

**f.** *val lys his.* If the order is *val lys his,* $val^+\ lys^-\ his^+$ should be rare, since this genotype results from a double exchange; and indeed, this class is the least frequent.

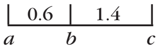
**27.** Mix the phages together with bacteria with increasing quantities of the two phages. Knowing the numbers of each in a particular case, it is possible to predict the proportion of cells doubly infected (product of probabilities). The recovery of recombinants should increase with that probability. In other words, recombination should occur only in doubly infected cells.

**29.** 

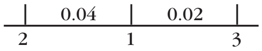
Small recombination frequencies should be approximately additive. Note that recombination distances are twice the value of wild-type plaques since the double mutant recombinants were not counted. Thus, the data table should be:

| Cross | Percent Wild-Type Plaques | Percent Recombinants |
|---|---|---|
| $a \times b$ | 0.3 | 0.6 |
| $a \times c$ | 1.0 | 2.0 |
| $a \times d$ | 0.4 | 0.8 |
| $b \times c$ | 0.7 | 1.4 |
| $b \times d$ | 0.1 | 0.2 |
| $c \times d$ | 0.6 | 1.2 |

The largest distance is between *a* and *c;* therefore *a* and *c* must be at opposite ends. Since *a-b* = 0.6, *b* must be 0.6 units to the right of *a.* This position gives *b-c* as 1.4, the observed distance. We now have the following map



If *d* is to the left of *b*, then *d-c* should be greater than 1.4, a result not seen. Therefore, *d* is 0.2 units to the right of *b.*

**31.** 

To calculate map distance, you must have the number of recombinations and the total number of progeny. Since all phages grow on strain B, this number must equal the total number of progeny; this is $250 \times 10^7$. Since only wild-type phages grow on K12, and since wild-types result from recombination between two genes,

$$\begin{array}{c}\underline{a \quad b^+}\\ \mathrm{X}\\ \overline{a^+ \quad b}\end{array} \longrightarrow \underline{a\ \ b} \text{ and } \underline{a^+\ \ b^+},$$

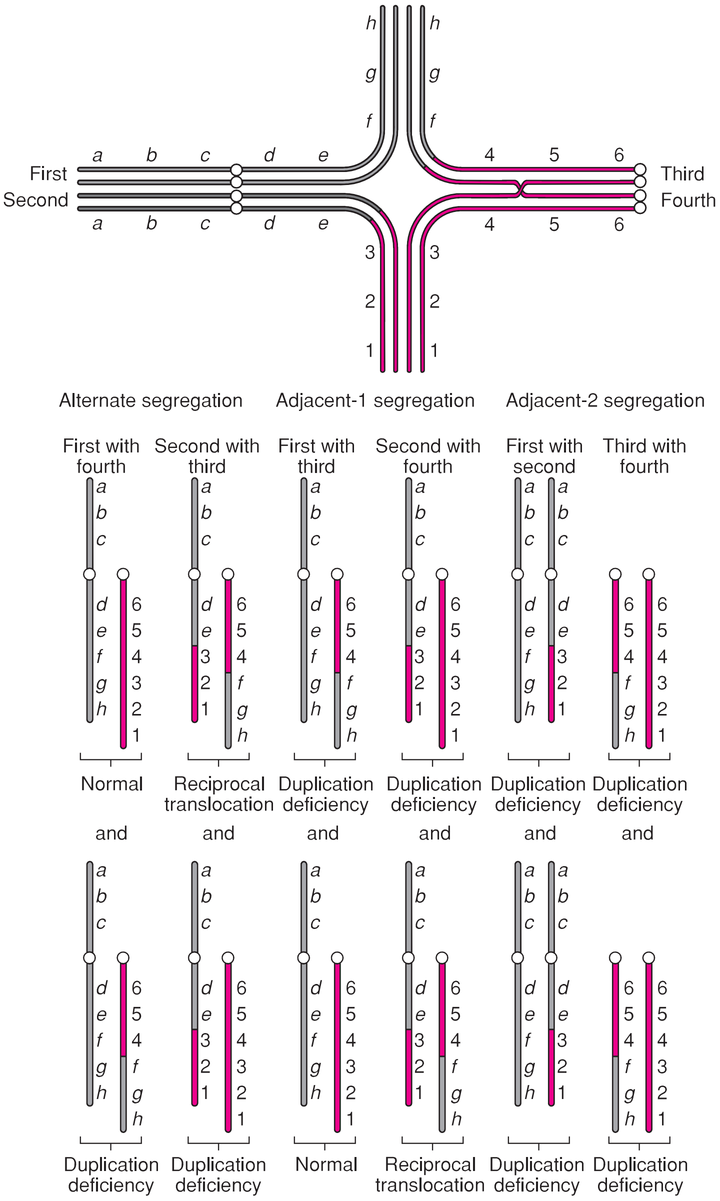
the number that grow on K12 must be recombinants. But this number represents only half of the recombinants, for the double mutant will not grow on K12. Total recombinants are:

$1 \times 2 \quad (2 \times 50)(10^4) = 10^6$
$1 \times 3 \quad (2 \times 25)(10^4) = 5 \times 10^5$
$2 \times 3 \quad (2 \times 75)(10^4) = 1.5 \times 10^6$

$$\textit{map distance } 1\text{–}2 = \frac{(100 \times 10^4)}{(250 \times 10^7)} \times 100$$
$$= (0.4 \times 10^{-3})(100) = 4 \times 10^{-2} = 0.04$$

$$\textit{map distance } 1\text{–}3 = \frac{(50 \times 10^4)}{(250 \times 10^7)} \times 100$$
$$= (0.2 \times 10^{-3})(100) = 2 \times 10^{-2} = 0.02$$

$$\textit{map distance } 2\text{–}3 = \frac{(150 \times 10^4)}{(250 \times 10^7)} \times 100$$
$$= (0.6 \times 10^{-3})(100) = 6 \times 10^{-2} = 0.06$$



33. Use replica-plating on selective media with arabinose as the sole carbon source, thus selecting for $ara^+$ cells. Although all three loci can be cotransduced, the rarity of $ara^+$ $leu^-$ $ilvH^+$ indicates *leu* is the middle locus *(ara leu ilvH).* Cotransductance frequencies:

$$\textit{ara} \text{ to } \textit{leu} = (9 + 340)/(9 + 340 + 32) = 0.92$$

$$\textit{ara} \text{ to } \textit{ilvH} = 340/(340 + 32 + 9) = 0.89$$

35. $azi^r$ $leu^+$ $thr^+$. The cotransduction frequency of *leu* and *azi* indicated *leu* is closer to *azi* than to *thr.* Thus, two orders are possible: *leu azi thr* or *azi leu thr.* If the first order is correct, and $leu^+$ is selected, $leu^+$ $azi^r > leu^+$ $thr^+$; this prediction fits. However, with the same gene order, and selecting for $thr^+$, $thr^+$ $azi^r > thr^+$ $leu^+$; this result is not seen, and the order must be *azi leu thr.* The second order predicts $thr^+$ $leu^+ > thr^+$ $azi^r$.

*Critical Thinking Question:*

1. At first, with much smaller data sets, scientists came up with branching models of the bacterial chromosome. However, with the very large data set of table 7.4, it is almost impossible to come up with a reasonable mechanism other than a circular bacterial chromosome.

## Chapter 8 Cytogenetics

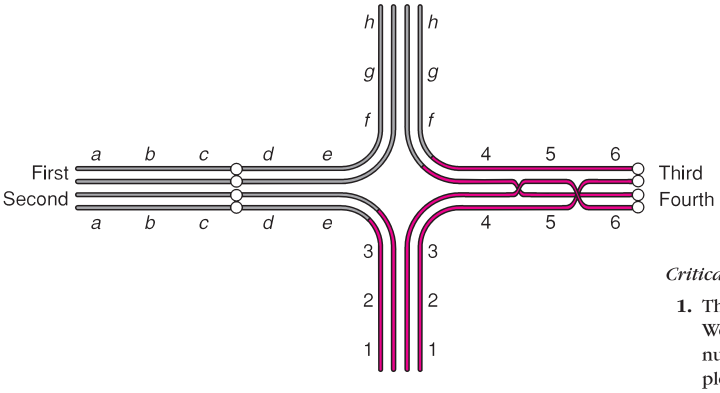
1. All chromosomes form linear bivalents. The cross-shaped figure is seen only in heterozygotes.
3. No, there are no inversion loops formed in homozygotes.
5. A diagram will show that a crossover between a centromere and the center of the cross can change the consequences of the pattern of centromere separation. For example, in the figure at left, we diagram a crossover between loci 4 and 5, as in figure 8.11.
7. Reciprocal translocation (some effects occur only in the heterozygous condition). Look for the cross-shaped figure at meiosis or in salivary gland chromosomes.
9. Assume crossovers as shown (following page, top left).
11. **a.** All females should get a wild-type X chromosome from their father. Irradiation produces chromosomal breaks, so a deletion of part of the X is possible, producing a situation of pseudodominance. (Alternatively, the offspring could have gotten a mutant $X^w$ from the father.)

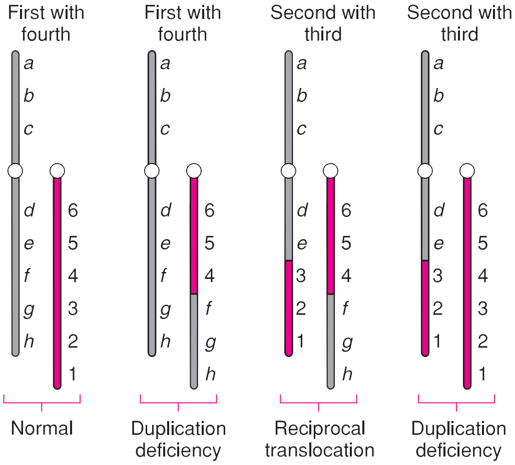
    **b.** Diagram the crosses (a "/" represents a deleted part of the chromosome and ________ = X chromosome)

    w__________ × +__________
    ////________ Y
    ↓

    | | | | |
    |---|---|---|---|
    | w______ | ////_____ | w____ | /////____ |
    | +______ | +_______ | Y | Y |
    | wild-type | wild-type | white | dead |

    We expect all wild-type females and all white-eyed males, but in a ratio of 2 female:1 male.
13. We see that the $F_2$ offspring from cross A × B yields fewer progeny than the other crosses. Something unusual must be involved. One explanation is that one of the strains is homozygous for a reciprocal translocation. The translocation, when heterozygous, results in some inviable gametes or progeny, and thus reduces the number of progeny.
15. We expect to see about 32% recombination between these two genes, but we see only 2%. The most likely explanation is that an inversion occurred so that these two genes came to lie close to each other; the stocks are homozygous. Since semisterility is not reported, we are probably not dealing with crossover suppression in inversion heterozygotes.





**17.** A translocation from the tip of the normal X in the male to the Y. We expect all males to receive an X chromosome with the white-eye allele from the female. For the male to be wild-type, we still must have part of the wild-type X chromosome. To test, cross this wild-type male with white-eyed females. All the female progeny should be white-eyed and all the male progeny red-eyed. Cytological examination of the chromosomes should reveal the translocation.

**19.** $4n = 92$; $2n - 1 = 45$.

**21.** $n = 8 + n = 6$ equals $14 \times 2 = 28$; $20 + 20 = 40$

**23.** An XO/XYY mosaic can occur by nondisjunction of the Y chromosome in a cell during early cleavage in an XY individual. An XX/XXY mosaic can come about if one of the cells during early cleavage in an XX zygote is fertilized again by a Y-bearing sperm. Trisomy 21 usually comes about from an egg with two copies of the chromosome; the egg had two copies because of meiotic nondisjunction.

**25.** The father. The allele for color blindness can only come from the mother. If meiosis in her is normal, an egg could get the X chromosome carrying the mutant allele. The daughter has only one X chromosome, so the sex chromosomes failed to separate in the man, and a sperm with neither X nor Y fertilized the egg.

**27.** The first meiotic division in the father is normal, producing cells with either two X or two Y chromatids. During the second meiotic division in the cell with the two Y chromatids, both Y chromatids move to the same pole and end up in the same sperm cell.

*Critical Thinking Question:*

**1.** The numbers are all multiples of 14 (1, 2, 3, 4, 5, 6, 7, and 8 copies). We can thus hypothesize that the original diploid chromosome number ($2n$) is 14. The other species would be polyploids, multiples of the original 14 (tetraploid, hexaploid, etc.). These are all the even ploids up to 112 chromosomes. As we saw, even ploids have the potential to succeed in meiosis, whereas odd ploids rarely do.

### Chapter 9 Chemistry of the Gene

**1.** The genetic code would somehow be read in number of tetranucleotide units, in which each unit consists of one each of the four bases (G, C, T, A). For example, one unit might be the amino acid alanine, two units might be the amino acid arginine, and so on.

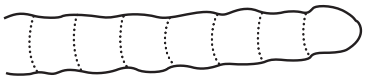
**3.** Sugars: DNA has deoxyribose, RNA has ribose; and bases: DNA has thymine in place of uracil, RNA has uracil in place of thymine.

**5.** See figures 9.18 and 9.19.

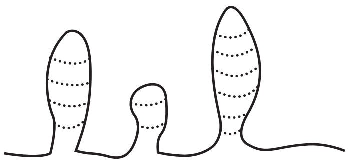
**7.** **a.** 28% G; 28% C; 22% A; 22% T. **b.** Same percentages except 22% U, 0% T. Chargaff's rule states that the quantity of A = T and the quantity of G = C. If G = 28%, then C = 28% and G + C = 56%. The sum of all bases must equal 100%. Therefore, (A + T) = 100 − 56 = 44. Since A = T, 1/2(44) = 22%. This is the amount of both A and T. For an RNA molecule, proceed the same way, except remember that U replaces T, so we have 22% U.

**9.** There must be regions of complementarity within the single-stranded regions. A melting temperature indicates some regions that are double-stranded. We can envision at least two different possible configurations:

1. Whole molecule complementary



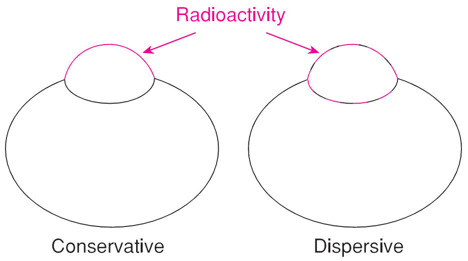
2. Fragment complementary



In fact, most single-stranded molecules have some regions that are complementary.

**11.** 19.1 billion base pairs. One base takes up 3.4 Å. Since 1Å is 1/10,000,000 millimeter, there are 10,000,000/3.4 = 2.94 million bases per millimeter. Multiply by 1,000 to get to meters, and finally by 6.5 for 6.5 meters: = $1.91 \times 10^{10}$ bases.

**13.**

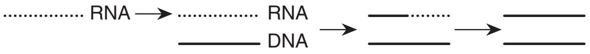


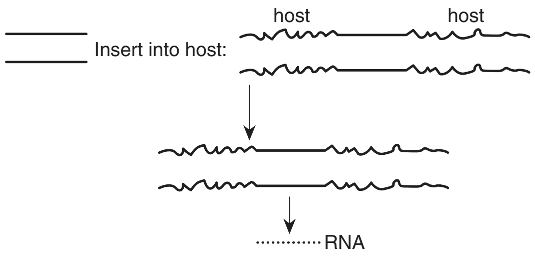
**15.** A primosome is a helicase plus a primase; it opens the DNA and creates RNA primers on lagging strands and is part of the replisome. A replisome includes a primosome plus two copies of DNA polymerase III; it coordinates replication on both the leading and lagging strands at the Y-junction.

**17.** See figure 9.28.

**19.** At one time molecular swivels, presumably protein in nature, located periodically along the DNA, were suggested.

**21.**





This question really asks how double-stranded DNA can be formed from single-stranded RNA. First, we could synthesize a complementary DNA strand, then begin to make a second DNA strand complementary to the first. While synthesizing the second DNA strand, we begin to degrade the RNA. The double-stranded DNA molecule now inserts into the host DNA. In order to get more viruses, a single-stranded RNA molecule must be made from the DNA. We would need, at a minimum, an enzyme to make a single-stranded DNA to form a hybrid with the RNA; an enzyme to degrade the RNA and to make the second, complementary DNA strand; an enzyme to cut the host DNA to allow the viral DNA to insert itself; an enzyme to ligate the two molecules; and an enzyme to make more viral RNA. The earlier functions are performed by one viral enzyme, reverse transcriptase.

**23.** Finding small pieces or fragments of DNA suggests the Okazaki pieces are only slowly, if at all, joined, a function of DNA ligase. The fact that not many long DNA molecules are seen also suggests that the DNA is being broken, implicating a nuclease as well.

**25.** It is unlikely that bases are added faster in developing embryos. So we must look for another mechanism. If there are more replicons, and hence more origins of DNA replication, each replicon will be shorter and be able to duplicate faster. Alternatively, and more likely, the process is regulated to slow down adult division.

*Critical Thinking Question:*

**1.** One way to study mutations that are generally lethal is by isolating temperature-sensitive mutations. These mutations involve amino acids that disrupt the functioning of the enzymes at some critical temperature but are phenotypically normal at other temperatures. Thus, the mutant organisms can be kept alive by growing them at one temperature (the *permissive* temperature) but their mutant effect can be studied at the temperature in which the protein function is disabled (the *restrictive* temperature). (For additional discussion of these mutations, see chapter 12.)

### Chapter 10 Gene Expression: Transcription

**1.** See figure 10.3. Complementarity is achieved between messenger RNA and ribosomal RNA and between messenger RNA and transfer RNA.

**3.** Transcription has higher error rates. Errors of DNA polymerase tend to become permanent, whereas errors of RNA polymerase do not.

**5.** A consensus sequence is made up of the nucleotides that appear in a significant proportion of cases when similar sequences are aligned. A conserved sequence consists of nucleotides found in all cases when similar sequences are aligned. For example, the Pribnow box (fig. 10.6) is the consensus sequence TATAAT.

**7.** See figure 10.8 for a promoter and figure 10.10 for a terminator. The transcript starting from the promoter would be 5′-CUUAUACGGU.... The transcript from the terminator is shown in figure 10.10.

**9.** A stem-loop structure can form when a single strand of DNA or RNA has a double helical section (see fig. 10.10). An inverted repeat is a sequence read outward on both strands of a double helix from a central point (see fig. 10.10). A tandem repeat is a segment of nucleic acid repeated consecutively; that is, the same sequence repeats in the same direction on the same strand:

5′-TCCGGTCCGGTCCGG-3′
3′-AGGCCAGGCCAGGCC-5′

A DNA sequence with a seven-base inverted repeat is

5′-ATTACCGCGGTAAT-3′
3′-TAATGGCGCCATTA-5′

**11.** Footprinting is a technique in which DNA in contact with a protein is exposed to nucleases; only DNA protected by the protein is undigested. Promoters could be isolated by protection with RNA polymerase in the absence of ribonucleotides—the polymerase will not move—and then sequenced.

**13.** The superscripts of the sigma factors refer to their molecular weights (e.g., $\sigma^{70}$ is 70,000 daltons). Different sigma factors usually recognize different prokaryotic promoters.

**15.** 3′ - G G T A G T A C T G T C T G G G A A C G A T T G C G -5′ ←
5′ - C C A T C A T G A C A G A C C C T T G C T A A C G C - 3′

Begin by writing the strand that is complementary to the RNA. This will be the transcribed strand. Remember, U in RNA pairs with A in DNA. Since transcription proceeds 5′ → 3′, the 5′ end of the RNA is opposite the 3′ end of the DNA.

**17.** The bottom strand is transcribed and the molecule is arranged as

5′ ________________3′
3′ ________________ 5′

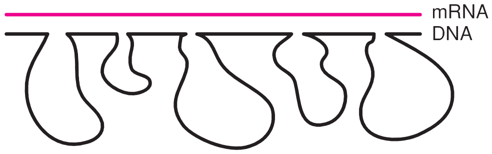
Begin by writing the RNA that could be transcribed from each strand. Since the DNA represents the beginning portion of the gene, the RNA must have an AUG to start protein synthesis. Unfortunately, both strands yield RNAs with one or more AUGs. The RNA from the top strand has AUGs in both directions, but in each case the AUG is followed by a termination signal, UAA, or UAG. This RNA could not make a protein. The bottom strand produces an RNA with only one AUG. Since transcription and protein synthesis both proceed 5′ → 3′, the left end of the bottom strand must be 3′.

**19.** If transcription of the genes is rho-dependent, the RNAs made at 40° C will be longer than those made at 30° C. Since the rho cannot function at the high temperature, RNA polymerase will read past the termination region. If transcription is rho-independent, a rho mutant will have no effect on transcription, and hence, the size of the RNAs.

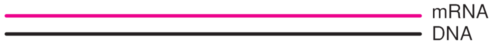
**21.** The double helix must unwind in order for transcription to occur. A-T pairs, because they have only two H-bonds, are more easily disrupted than G-C pairs.

**23.** Removing one base too many or too few would result in a shift in the reading frame during translation (see chapter 11), thus radically altering the protein product.

**25.** Five introns



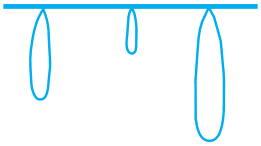
No introns



**27.** Group I introns are self-splicing introns that require a guanine-containing nucleotide for splicing. Group II introns are similar but do not require an external nucleotide for splicing. Group I and II introns are released as linear and lariat-shaped molecules, respectively.

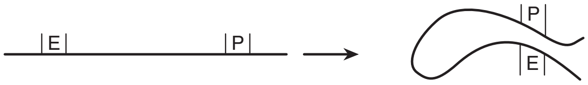
**29.** Spliceosomes are composed of small nuclear ribonucleoproteins (see fig. 10.37).

**31.**



There are three introns, so we expect three single-stranded DNA loops. The coding regions (exons) form RNA-DNA hybrids and appear thicker.

**33.** Enhancers [E] bind activator proteins that also bind proteins of the polymerase at the promoter [P]. In some eukaryotic genes, there are many enhancers, allowing numerous levels of control and forcing enhancers further and further upstream. For activators bound to enhancers to bind polymerase proteins, the DNA must loop around.



*Critical Thinking Question:*

**1.** Given that a gene controls the production of a protein, there are both realistic and theoretical limits to the size of a gene. From what we know about the functioning of the centromere, a gene would have to occupy no more than the length of a chromosomal arm. However, given that human chromosomes must contain about fifty thousand genes, it is unlikely that any gene is that large. We also know that many functioning proteins are made up of subunits, each controlled by its own gene. Thus, large proteins tend to be conglomerates of smaller ones rather than large functional units. Finally, the larger the protein the more time it takes to transcribe and translate it, making very large size inefficient. As we mentioned before, the average protein is about 300 to 500 amino acids; with introns, and control elements, the gene for an average protein could be quite large. The largest known gene is the human dystrophin gene that codes for a cytoskeletal protein. It is 2,300,000 bases long, has 79 exons, and takes 16 hours to be transcribed.

### Chapter 11 Gene Expression: Translation

**1.** The messenger RNA is 5′-AUGUUACCGGGAAAAUAG-3′; the anticodons are 3′-UAC-5′, 3′-AAU-5′, 3′-GGC-5′, 3′-CCU-5′, 3′-UUU-5′; the amino acids are methionine, leucine, proline, glycine, lysine (see the figure below).

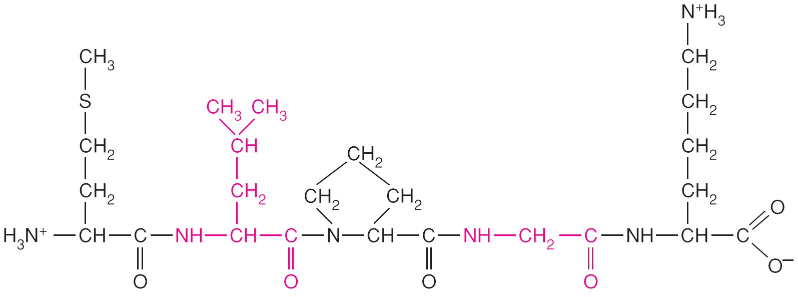
**3.** See figure 11.16. Use the messenger RNA of problem 1 and be sure to include EF-Tu and EF-Ts.

**5.** There are approximately twenty aminoacyl-tRNA synthetases in an *E. coli* cell, one for each amino acid. Recognition signals can occur at any point on a given transfer RNA, although the anticodon figures prominently in most.

**7.** See figure 11.7.

**9.** Three; see figure 11.29.

**11.** 5′-UAA-3′ → 5′-UUA-3′ (leucine). The consequence is the failure to terminate the particular protein leading to continued chain elongation to the next nonsense codon or to the end of the messenger RNA. The result is probably a nonfunctioning enzyme or protein.



**13.** EF-Tu brings a charged transfer RNA to the A site at the ribosome. EF-Ts is involved in recharging EF-Tu (see fig. 11.14). The eukaryotic equivalents are eEF1$\alpha$ and eEF1$\beta\gamma$.

**15.** A signal peptide is a sequence of amino acids at the amino-terminal end of a protein that signals that the protein should enter a membrane (see fig. 11.25). Although the concept is the same, the situation in eukaryotes is somewhat more complex because there are many different membrane-bound organelles, each having their own membrane-specific requirements. Signal peptides are usually cleaved off the protein after the protein enters or passes through the membrane.

**17.** $NH_2$-FGKICABHLNOEDJM-COOH

**19.** **a.** 5′-AUG AUU GAA UGC GAG CGG AGU-3′

**b.** N-met-ile-glu-cys-glu-arg-ser

First determine the sequence of the RNA complementary to the given DNA strand. Don't forget about polarity; as the strand is written, the 5′ end of the RNA will be on the left. Blocking off successive groups of three bases allows the determination of the codons. Use the code to determine the amino acid sequence.

**21.** 12/27 phenylalanine (UUU, $[2/3]^3$; UUC, $[2/3]^2[1/3]$); 6/27 serine (UCU, $[2/3]^2[1/3]$; UCC, $[2/3][1/3]^2$); 6/27 leucine (CUU, $[2/3]^2[1/3]$; CUC, $[2/3][1/3]^2$); 3/27 proline (CCU, $[2/3][1/3]^2$; CCC, $[1/3]^3$).

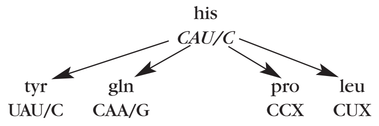
**23.** The table could look the same (see table 11.4) except that the position would be left side = first position (5′ end); top = third position (3′ end); right side = second position. For example, the codons for valine (currently 5′-GUU-3′, 5′-GUC-3′, 5′-GUA-3′, and 5′-GUG-3′) would be 5′-GUU-3′, 5′-GCU-3′, 5′-GAU-3′, and 5′-GGU-3′.

**25.** We are mixing two RNA strands that are complementary; these strands will form a double-stranded RNA molecule. Since we observed the incorporation of no amino acids, the ribosome must not be able to read a double-stranded molecule.

**27.** If we write out $(GUA)_n$ as GUA GUA GUA GUA . . ., we see that we could use any of three different reading frames: GUA, UAG, or AGU. Since we see only two amino acids incorporated, either two of the possible codons code for the same amino acid, or one of the codons is a stop codon. If you look at the code, you will see that UAG is a stop codon.

**29.** The stop codon has probably mutated to give a codon for the amino acid leucine. The longer-than-normal protein suggests that the original stop was not read. Numerous possibilities exist. If the second letter of a stop codon were changed to a U, we would have UUA or UUG leucine codons. Alternatively, the insertion of a C before the U would yield CUA, CUG, or CUA as leucine codons. Similarly, an insertion of a U next to the first G yields UUA or GUG as leucine codons. Since the next amino acid is phenylalanine, the next codon must be UUC or UUU. If a base were added, as above, the next codon would have to begin with A or G, and phenylalanine does not begin with A or G. Therefore, the most likely explanation is a change of the second letter from an A to a U.

**31.** Either CAU or CAC. Write down all possible codons for each amino acid.



For *leu,* note that UUA/G cannot result from a single change in the *his* codon. Therefore, leucine must be CUX. All of the other amino acids could result from single changes in either the first or second base, and we are left with either codon being the one for *his.*

*Critical Thinking Question:*

**1.** In general, transcriptional and translational signals are independent. We could look at this by asking the question, how does changing one of the signals affect the other process? In other words, if we changed a translational signal such as a start or stop codon, would that affect the transcription of that gene? In general, the answer is no.

### Chapter 12 DNA: Its Mutation, Repair, and Recombination

**1.** For example, if the twenty individual cultures of table 12.1 had values of 15, 13, 15, 20, 17, 14, 21, 19, 16, 13, 27, 14, 15, 26, 12, 21, 14, 17, 12, 14, then the mutation theory would not have been supported because the variation between the individual and bulk cultures would not have been different.

**3.** Reading across each row, we gather more and more information.

Row 1: 1, 6, and 7 are part of one complementation group.

Row 2: 2 and 5 are part of one complementation group.

Row 3: 3 and 4 are part of one complementation group.

Row 4: no new information.

Row 5: no new information.

Row 6: reinforces that 6 and 7 are part of the same complementation group.

Row 7: no new information.

Thus, we conclude that there are three complementation groups present: 1, 6, and 7 are mutually noncomplementing, as are 2 and 5, and 3 and 4. The half-table missing is a mirror image because a cross of 1 and 3 is the same as a cross of 3 and 1 (reciprocity). The diagonal always contains negative elements because every mutant is a functional allele of itself.

**5.** All mutants should be crossed in pairwise combinations yielding heterozygote daughters. (Presumably, earlier crosses indicated that these are X-linked loci.) All $F_1$ daughters will be wild-type: The mutations complement and therefore are not alleles. When eosin flies are crossed in a similar fashion, daughters will be wild-type except when the parents were eosin and white. In that case, daughters will be mutant, showing the lack of complementation and hence that the mutations are alleles.

**7.** Prokaryotic and phage genes generally do not have intervening sequences. Benzer and Yanofsky worked at a time when introns were unknown and it was assumed that the length of a gene was transcribed and then translated. If the genes had introns, Benzer and Yanofsky would have been unaware of them; introns would not affect colinearity or mapping. Introns would affect physical measures of the lengths of DNA, with which Benzer and Yanofsky were not involved.

**9.** The *rex* gene of phage $\lambda$ represses growth of phage T4 *r*II mutants.

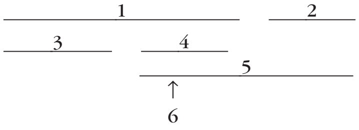
**11.** In replicating DNA, a transition mutation can occur by tautomerization of a base in the template strand (template transition) or entering the progeny strand (substrate transition).

**13.** 5-bromouracil (pyrimidine analogue) and 2-aminopurine (purine analogue) are incorporated into DNA as thymine and adenine, respectively. However, each undergoes tautomeric shifts more frequently than the normal base. Both cause transitions. Nitrous acid also promotes transitions by converting cytosine into uracil, which acts like thymine, and adenine into hypoxanthine, which acts like guanine. Proflavin induces insertions and deletions by intercalating and buckling DNA. Ethyl ethane sulfonate removes purine rings and thus promotes transitions and transversions.

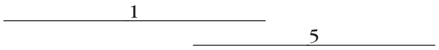
**15.** See figure 12.25.

**17.** Three genes. Gene *A:* mutants 1, 4, 8; gene *B:* mutants 2, 5; gene *C:* mutants 6, 7. Mutant 3 probably contains a deletion that spans genes *A* and *C.* Begin by finding mutants that do not complement. These should have mutations in the same gene. Mutants 2 and 5 are in the same gene. Initially, we suspect that mutants 1, 3, 4, and 8 are in the same gene, which is different from the gene that contains 6 and 7. If mutant 3 is in gene *A,* it should complement 6 and 7, and it does not. One explanation is that 3 is a deletion spanning genes *A* and *C.* Alternatively, mutant 3 could be in gene *A* but be a polar mutation. Either possibility implies that the order is *B A C.*

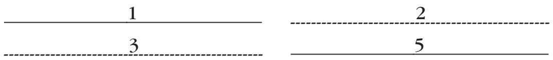
**19.**



Begin with deletions that yield mostly "−"s. These must be large deletions that cover most of the other deletions. Mutations 1 and 5 are such mutations. Since they give no wild-type, they must overlap:



Now look at mutant 2. It gives wild-type recombinants with 1 but not 5. Therefore it must overlap the region deleted in 5. Mutant 3, by similar logic, must cover part of deletion 1. We can draw these results as follows. Broken lines indicate we do not know yet how long the deletion is:



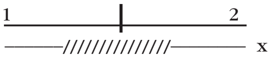
Now look at 4. It gives "−" (no wild-type recombinants) with both 1 and 5 and therefore must be in the region that 1 and 5 overlap. If 4 extended to and overlapped either 2 or 3, we expect to see "−" with them. Since this prediction is not met, 4 must be a small deletion spanning at least part of the overlap of 1 and 5. Since mutant 6 gives no wild-type with 1, 4, or 5, it must be within the common region deleted in all three strains.

**21.** We know that the anticodon pairs with the codon, and we expect nonsense suppressors to contain an altered anticodon. The fact that the nonsense codon can be read by a transfer RNA with a normal anticodon but altered dihydrouridine loop suggests that the way in which this loop interacts with the ribosome causes the anticodon sequence to be misread.

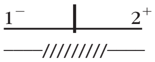
**23.** $x^+ \rightarrow x$: AT → GC. $y^+ \rightarrow y$: GC → AT. This problem requires logic and a knowledge of how mutagens work. For *x,* the key is the response to HA, which only causes GC → AT transitions. Mutant *x* is reverted by HA, therefore, *x* must be GC, and the normal $x^+$ was AT. AP-induced mutations can also be reverted by AP. Since *y* is not reverted by HA, *y* must be AT. Therefore, $y^+$ must be GC.

**25.** A deletion that spans regions of more than one gene. Consider the following two genes:

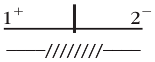
By definition, mutations in gene 1 will complement mutations in gene 2 but not other mutations in gene 1. If we have a deletion, x, that covers part of both 1 and 2 (slashes),



presumably gene 1 will be nonfunctional because it is missing the last part of the protein. Gene 2 will be nonfunctional because the beginning portion of the gene is missing. The following genotypes will give no complementation:



No functional product of gene 1



No functional product of gene 2

We could also get a lack of complementation when mutants are in two genes if we have a bacterial operon in which one of the genes contains a polar mutation creating a transcription stop signal. Such a mutation eliminates all distal functions. Thus, if the operon is

$$\underline{A \quad B \quad C \quad D},$$

and we construct the partial diploid

$$\frac{A^-(\text{polar}) \quad B^+C^+D^+}{A^+ \quad B^-C^+D^+},$$

we will get no complementation because the top DNA is effectively $A^-\ B^-\ C^-\ D^-$.

**27.** The auxotroph probably contains a deletion. If a few bases are missing, nothing is available to cause transitions or transversions. It is highly unlikely that the correct number of missing bases could be spontaneously and correctly inserted.

**29.** Excision repair endonucleases can recognize dimerizations, mismatched bases, and apurinic-apyrimidinic sites.

**31.** See figure 12.38 for a diagram of recombination. Branch migration is shown in figure 12.39.

*Critical Thinking Question:*

**1.** The gene is a linear entity that specifies the linear order of amino acids in a protein in a colinear fashion. Although scientists in the 1960s were convinced of colinearity, there were other alternatives possible. For example, DNA could be a branching structure. Or, transcription of DNA could take place such that the beginning, middle, and end of the gene were not in order. Thus colinearity supported our understanding of the shape and functioning of DNA.

### Chapter 13 Genomics, Biotechnology, and Recombinant DNA

**1.** Type II endonucleases are valuable because they cut DNA at specific points and many leave overlapping or "sticky" ends.

**3.** In DNA with a random sequence, a four-cutter will find sites approximately once in $4^4$ bases (= 1/256 = 0.0039). A six-cutter will find sites approximately once in $4^6$ bases (= 1/4096 = 0.0002). An eight-cutter will find sites approximately once in $4^8$ bases (= 1/65,536 = 0.000015).

**5.** DNA can be joined by having compatible ends to begin with or by blunt-end ligation (linkers combine these methods). The appropriateness of a method depends on what DNA is to be cloned and how that DNA can be obtained. Having DNA with "sticky" ends created by the same restriction enzyme would be easiest but sometimes is not available. Adding linkers by blunt-end ligation with a particular restriction site is usually the best compromise.

**7.** A plasmid is a self-replicating circle of DNA found in many cells. Foreign DNA inserted into a vector forms an expression vector if that foreign DNA produces a protein product. Cosmids are plasmids that contain *cos* sites and are useful for cloning large segments of DNA (up to 50 kb). YACs, yeast artificial chromosomes, have the loci to replicate in yeast (centromere, replication origin, and telomeres). They can be used to clone very large pieces of DNA, upwards of one million bases.

**9.** Chromosome walking is a technique for cloning overlapping chromosomal regions starting from an arbitrary point (see fig. 13.38). It is useful for determining relative locations of genes in uncharted regions as well as cloning regions too big to fit in a single vector.

**11.** Southern and northern blotting are gel transfer techniques used to probe for DNA and RNA sequences, respectively. Western blotting is a technique used for locating a protein by antibody recognition. Dot blotting is a probing technique for cloned DNA that eliminates the electrophoretic separation step.

**13.** Plasmids of *E. coli* origin survive in yeast when a yeast centromere (CEN region) is added, allowing them to replicate within the yeast cell. Inactivated SV40 viruses can function in the presence of intact "helper" viruses that allow them to complete their life cycles. Phage λ has parts of its chromosome that can be removed while still allowing it to complete its life cycle.

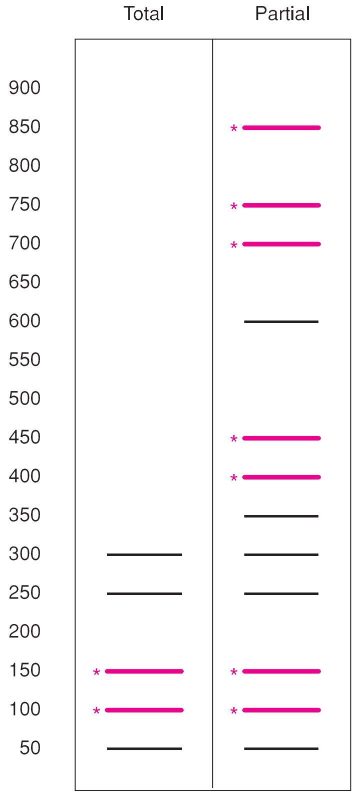
**15.** Partial digestion of molecule 2 leads to the following molecule:

```
AAAAAAAA
     TTTTTTTT
```

Some of these molecules will form a circle with the single-stranded Ts paired with single-stranded As. The circle eliminates the free 5′ phosphate, and the enzyme can no longer work.

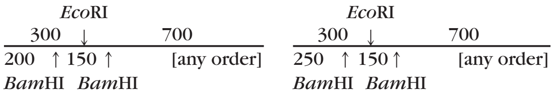
**17.** We must insert DNA that has no introns into bacterial plasmids. This DNA can be obtained by isolating mature, cytoplasmic messenger RNA and then using reverse transcriptase to make double-stranded cDNA. Plasmids with cDNA inserted can then be used to produce human proteins (expression vectors).

**19.** Electrophoretic bands of the total digest are (* indicates end label) 50, 100*, 150*, 250, 300 bp. Bands of the partial digest are 50, 100*, 150*, 250, 300(×2), 350, 400*(×2), 450*(×2), 600, 700*, 750*, 850* bp.



**21.** Mutant A: elimination of site between the 300- and 50-bp segments. Mutant B: elimination of site between 100- and 300-bp segments. Mutant C: creation of a new site within the 300-bp segment, dividing it into 75- and 225-bp segments.

**23.** This problem begins as a trial and error attempt to overlay two restriction maps, made more difficult by the fact that one enzyme, *Bam*HI, has made three cuts that are unordered, leaving many possibilities. However, a bit of thought beforehand makes this problem much easier. If you compare the double digest with the *Bam*HI digest, they share 200-, 250-, and 400-bp segments. The double digest has 50- and 100-bp segments replacing the 150-bp segment in the *Bam*HI digest. The inference is that there is an *Eco*RI cut in the 150-bp segment leading to the 50- and 100-bp segments, with all other segments of the *Bam*HI digestion left uncut. That leaves only two possibilities, as shown below; the data are insufficient to distinguish between the two choices.



**25.**

*Eco*RI 6.2 | 2.8 | 4.6 | 7.4 | 8.0

*Bam*HI 10.0 | 13.0 | 6.0

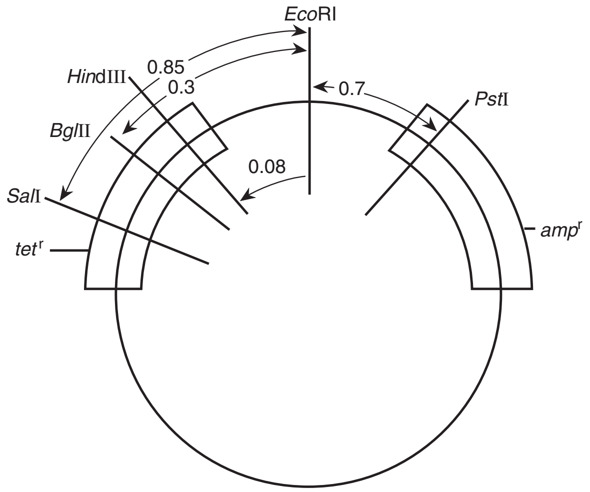
We know that the 6.2 and 8.0 kb *Eco*RI fragments are at opposite ends, and that the 10.0 and 6.0 kb *Bam*HI fragments are at the ends. Therefore, the *Bam*HI 13.0 kb fragment must be in the middle. If the 6.2 and 6.0 kb fragments are at the same end, a double digest should produce a fragment of 0.2 kb. This is not seen, so they are at opposite ends:

6.2 | (2.8, 4.6, 7.4) | 8.0

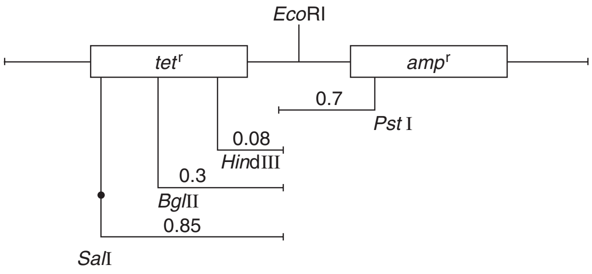
10.0 | 13.0 | 6.0

If 7.4 is next to 6.2, we should see a 3.8 fragment. Similarly, if 4.6 is next to 6.2, we should see a 3.8 fragment. We see neither of these fragments, so 2.8 is next to 6.2. The 4.6 fragment must be next.

**27.** Since *Eco*RI does not eliminate either resistance, its site must be between the *tet*$^r$ and *amp*$^r$ genes. The *Pst*I site must be within the *amp*$^r$ gene, since insertion of DNA into this site eliminates ampicillin resistance. By similar logic, the other sites must be in the *tet*$^r$ gene. In the double digests, the smaller fragment must represent the distance from the *Eco* RI site to the other site. We can draw part of the plasmid as:



(a)



(b)

If this arrangement is correct, digestion with *Bgl*II, *Eco*RI, and *Pst*I should yield 0.3 kb (*Bgl*II + *Eco*RI), 0.7 kb (*Eco*RI + *Pst*I) + 2.0 kb (*Pst* I + *Bgl*II), and this is, in fact, observed.

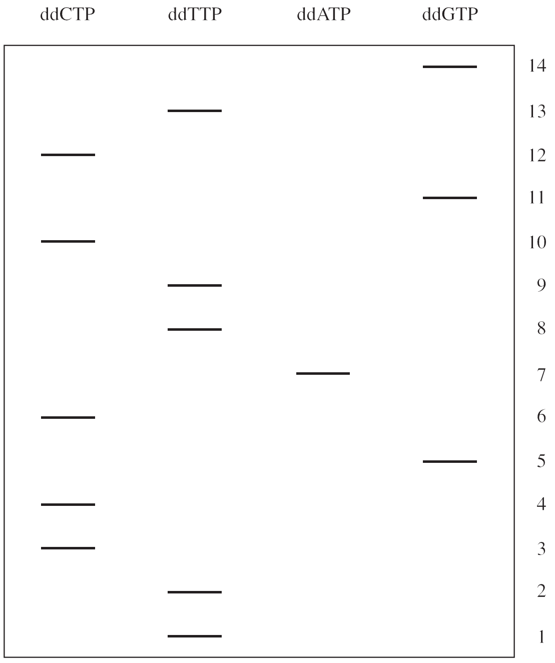
**29.** The second possibility predicts that a 1.5-kb fragment should be seen in the double digestion. This fragment would result from digestion of the 3.5-kb *Eco*RI fragment with *Bgl*II. Since we don't see a 1.5-kb fragment in the double digest, the second possibility does not agree with the observed results.

**31.** Two of the three number 21 chromosomes present in the child came from the father, not the mother. Since the probe produces different bands in the mother and the father, all of these bands must also be present in the child. The intensity of a band is proportional to the amount of DNA present. The bands that are of paternal origin are more intense than the maternal bands. The father contributed two number 21 chromosomes.

**33.** For the steps in the dideoxy sequencing method, see figures 13.32 and 13.33. Use of fluorescent dyes has allowed for the automation of the process and the elimination of radioactive tags.

**35.** The DNA can be inserted into the M13 general sequencing vector.

**37.** What will appear in the gel are fragments of the newly synthesized strand. Since DNA synthesis proceeds 5′ → 3′, the 5′ base will be T in the new strand. Proceed up the gel by indicating the base complementary to the sequence given.



**39.** Hypervariable DNA is DNA showing a great deal of interindividual variation. A RFLP (restriction fragment length polymorphism) is a polymorphism (variation) that shows up after Southern blotting and probing of restriction digests. A VNTR (variable number of tandem repeats) locus is one that is hypervariable due to unequal crossing over among the tandem repeats. VNTR loci are a hypervariable subset of RFLPs. Sequence-tagged sites are unique sites in the genome that can be amplified with polymerase chain reaction. Microsatellite DNA, repeats of very short segments such as CA, forms VNTR loci that are usually examined by polymerase chain reaction (PCR) if primer sequences are known.

*Critical Thinking Question:*

**1.** Sticky ends would exist if the plasmid DNA had a 3′ overhang of one nucleotide residue, while the foreign DNA had a 3′ overhang of its complement. The exact number of bases would not need to be the same because repair enzymes plus ligase could close the gap. For example, the plasmid could have a 3′ tail of thymines added, whereas the foreign DNA could have a 3′ tail of adenines added. This method (see figure next page) was called the poly-dA/poly-dT technique.

### Chapter 14 Gene Expression: Control in Prokaryotes and Phages

**1.** **a.** inducible (wild-type); **b–d.** constitutive; **e.** neither, super-repressed; **f.** inducible

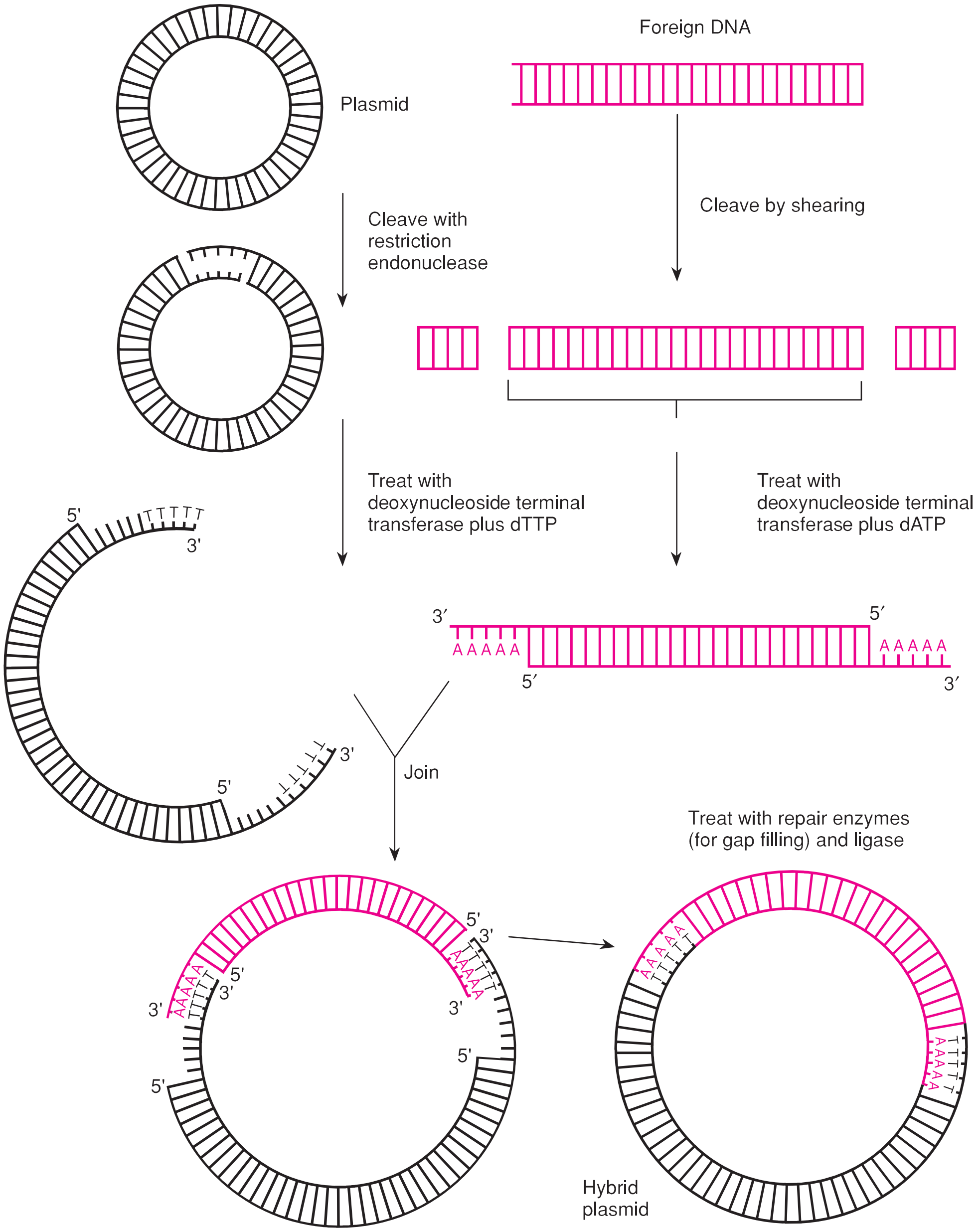
**3.** One mutant could fail to bind to operator DNA but could still bind the inducer ($op^-$, $in^+$). This mutant would have constitutive transcription of the operon. The reverse situation could also be true; the repressor could bind the operator but not the inducer ($op^+$, $in^-$). This mutant would be off all the time. A third mutant could fail to bind both ($op^-$, $in^-$), being constitutive.

**5.** **a.** Operator and repressor. **b.** Make a partial diploid with wild-type; the operator mutant will make β-galactosidase constitutively, and the repressor mutant will make it only in the presence of lactose. Four mutations are possible in the *lac* operon: mutations in the *z* gene or the promoter never make the enzyme. Mutations in the repressor always make the enzyme because the repressor cannot bind DNA. In operator mutations, a good repressor can never bind DNA. In a partial diploid, $i^-o^+/i^+o^+$, the wild-type repressor is *trans* acting and can bind to both operators, creating an inducible situation. In $i^+o^-/i^+o^+$, repressor cannot bind to $o^-$, and this DNA is always on, even though the wild-type DNA is off in the absence of lactose.

**7.** Cyclic AMP, combined with CAP protein, attaches to CAP sites enhancing transcription of nonglucose, sugar-metabolizing operons in *E. coli.* Glucose inhibits its formation by inhibiting adenylcyclase.

**9.** We must think about how these operons are controlled. Not only do they need inducer, but they also require the catabolite repression-activation system. These mutants could be unable to make cAMP because the adenylcyclase gene is defective. Alternatively, they could be making a defective catabolite activating protein (CAP).

**11.** $b$ = tryptophan synthetase gene, $a$ = operator, $c$ = repressor. Look first for the single mutation that never gives enzyme; this genotype will give the letter of the structural gene. The genotype $a^+b^-c^+$ fits this requirement, so $b$ is the structural gene, and $a$ and $c$ represent control regions. Genotypes 4 and 5 tell us nothing. Look at genotype 6. The right DNA will never make the enzyme. If $c$ is the operator, the left DNA should always make the enzyme, and this is not seen. Therefore, $a$ must be the operator. Check these assignments with genotype 7. The right DNA will never make the enzyme. If $a$ is the operator, the left DNA is always on. If $a$ is the repressor, the right DNA makes a good repressor that will bind to the top DNA and regulate it.


Foreign DNA
Plasmid
Cleave with restriction endonuclease
Cleave by shearing
Treat with deoxynucleoside terminal transferase plus dTTP
Treat with deoxynucleoside terminal transferase plus dATP
5'
TTTTT
3'
3′
5′
AAAAA
AAAAA
5′
3′
3'
5'
TTTTT
Join
Treat with repair enzymes (for gap filling) and ligase
5'
3'
AAAAA
TTTTT
3'
5'
3'
AAAAA
TTTTT
3'
5'
5'
AAAAA
TTTTT
TTTTT
AAAAA
Hybrid plasmid

**13.** The *E. coli trp* operon functions as a normal repressible operon. In addition, attenuation, based on secondary structure and stalling of the ribosome on the leader transcript, can further prevent transcription (see fig. 14.16). Attenuator control can be exerted based on other amino acids if their codons appear in the leader transcript, causing ribosome stalling.

**15.** Assuming that the mutants produce inactive proteins: *cI, cII,* and *cIII,* lytic response; *N,* neither lytic nor lysogenic responses possible; *cro,* no lytic response possible; *att,* no lysogenic response possible; *Q,* no lytic response possible.

**17.** The λ chromosome has one circular and two linear forms. The circular form is the infective cellular form. A break at one point (*cos* site) takes place during packaging into the phage head, and a break at another point forms the linear integrative prophage (see fig. 14.18).

**19.** The prophage region of the Hfr chromosome enters the $F^-$ cell with no repressor present. The situation is thus similar to regular phage infection, which can go either way (lysogenic or lytic cycles).

**21.** The cells that form colonies do not contain a prophage. The initial heat shock inactivated the repressor and allowed the prophage to excise. The lower temperature activated the repressor, and the repressor bound to the excised phage DNA, preventing gene expression and reintegration of some of the phages. These "cytoplasmic" phages failed to replicate and were lost during cell division. The cells that retained the virus were lysed when the temperature was raised because the phage could now make RNA.

**23.** An IS element is a simple transposon, which is a segment of DNA that can make a copy of itself to be inserted at another place in the genome. An intron is an intervening sequence, a region excised from messenger RNA before expression. A plasmid is an autonomous, self-replicating genetic particle. A cointegrate is an intermediate structure in transposition.

**25.** See figure 14.29.

**27.** See figure 14.35.

**29.** Transcription is the level at which most control mechanisms work. These include sigma factors, efficiency of promoter recognition, catabolite repression, operon-repressor systems, attenuation, and transposition. Translational control mechanisms include polarity placement, antisense RNA, differences in the efficiency of processes due to nucleotide sequence differences, codon preference, and the stringent response. Posttranslational mechanisms include feedback inhibition and differential rates of protein degradation.

**31.** Heat shock proteins are normally induced by the presence of a specific sigma factor, which itself is induced by heat shock.

**33.** The stringent response is the response of a prokaryotic cell to amino acid starvation. The idling reaction of the ribosome results in production of 3′-ppGpp-5′, whose appearance is associated with the cessation of transcription, especially of transfer RNAs and ribosomal RNAs, through an unknown mechanism.

**35.** Feedback inhibition: allosteric enzymes in some synthetic pathways can be inhibited by the end product of that pathway. In addition, many repressor proteins are allosteric.

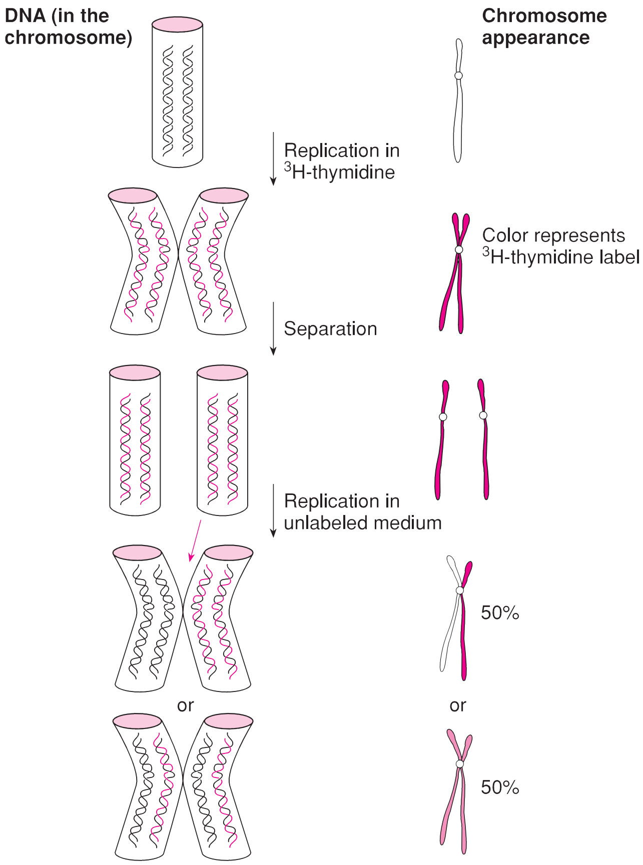
*Critical Thinking Question:*

**1.** One might think that a bacterium should metabolize any sugar in its environment. However, glucose is the most efficient sugar to metabolize and if it is present, it should be metabolized first. Even under maximal growth, it takes an *E. coli* cell about 20 minutes to grow and divide. If the cell were to begin to take up and metabolize other sugars, there would be a cost for the manufacture of new enzymes and the inefficiency of the initial steps of the sugar metabolism. In fact, the growth of the cell would slow down under these circumstances and be at an evolutionary disadvantage.

### Chapter 15 The Eukaryotic Chromosome

**1.** In general, prokaryotes are small, have a relatively small circular chromosome, and have little internal cellular structure compared to eukaryotes. Most prokaryotic messenger RNAs are polycistronic, under operon control; eukaryotic messenger RNAs are highly processed, monocistronic, and usually not under operon control. Prokaryotes are mostly single-celled organisms, whereas eukaryotes are mostly multicellular. Eukaryotes have repetitive DNA, absent for the most part in prokaryotes. Prokaryotic chromosomes are not complexed with protein to anywhere near the same extent that eukaryotic chromosomes are.

**3.** Assume that each chromosome contained two complete copies of the same DNA. Following the protocol of figure 15.1, the final results would be chromosomes, before separation, that consisted of either two labeled chromatids or only one labeled chromatid, in a 1:1 ratio (barring sister chromatid exchanges). The labeled chromatid in chromosomes with just one chromatid labeled will have twice the label of each chromatid in the chromosomes in which both chromatids are labeled (see the following figure).


DNA (in the chromosome)
Chromosome appearance
Replication in $^3$H-thymidine
Color represents $^3$H-thymidine label
Separation
Replication in unlabeled medium
50%
or
or
50%

**5.** The length of DNA associated with nucleosomes was determined by footprinting, in which free DNA was digested, leaving only those segments protected by nucleosomes. Nucleosome hypersensitive sites are sites not in a nucleosomal state; they seem to be sites involved in the initiation of replication, transcription, and other DNA activities.

**7.** See figures 15.10 and 15.11 for the relationship of the 110, 300, and 2,400 Å chromosome fibers.

**9.** See figure 15.19.

**11.** Polytene chromosomes are chromosomes that underwent endomitosis: They consist of numerous copies of the same chromatid (e.g., in the *Drosophila* salivary glands). Regions of active transcription in polytene chromosomes form diffuse areas called puffs or Balbiani rings (see fig. 15.14). Lampbrush chromosomes occur in amphibian oocytes (see fig. 15.18).

**13.** Satellite DNA differs in its base sequence from the main quantity of DNA and thus forms a satellite band during buoyant density analysis. It is usually centromeric heterochromatin, composed of a highly repetitive DNA.

**15.** Telomeres are repetitive DNA sequences at chromosomal ends. They are repetitions of a five- to eight-base sequence. Most telomeres are G-rich. Telomeres protect the ends of chromosomes and probably provide signals on senescence of cells.

**17.** Highly repetitive DNA usually makes up the centromeric and telomeric regions of the chromosome. Unique DNA, making up the bulk of structural genes, has a large component that is transcribed. Repetitive DNA is composed of dispersed DNA (e.g., short and long interspersed elements—SINES and LINES), multiple copies of transcribed DNA (e.g., ribosomal RNA, histones), and diverged copies of ancestral genes (e.g., globin family genes).

**19.** The most direct method of determining the direction of transcription of the histone genes would be to clone and sequence the region, from which transcriptional information can be ascertained.

**21.** Cloning and then sequencing the region would provide the answer. Analysis would show genes of similar sequence to the active genes but lacking the sequences for transcription.

**23.** It is very rich in A-T sequences. Density is proportional to G-C content. Since the molecule has a low density, it has low G-C, and therefore high A-T.

**25.** Spaces between the nucleosomes must contain many promoter sequences. For the DNA to be digested, it must be unprotected. Since we see little transcription, the promoters must be missing and must have been destroyed by the nucleases.

**27.** Highly repetitive DNA must be located in these regions. Since most highly repetitive DNA is not transcribed, the results suggest that centromeric and telomeric regions are not transcribed.

**29.** The C-value paradox involves the issues of the excessive amounts of DNA in eukaryotic cells and the difference between eukaryotic species that seem to have similar complexity. It is explained by the large amount of structural DNA in chromosomes as well as the large amounts of short and long interspersed elements (SINEs and LINEs).

*Critical Thinking Question:*

**1.** Comparative DNA studies can be helpful in understanding the roles of the various types of DNA in the eukaryotic chromosomes if there are cases in which there are remarkably large differences in the amount of DNA in similar species. It can then be inferred that the basic developmental plan of an organism is contained in the one with the lower amount of DNA, and the extra DNA in the species with more DNA may be superfluous. We do have cases in which amphibians differ by as much as one hundred times the amount of DNA found in similar species. The puffer fish has only one-sixth the amount of DNA as other higher eukaryotes.

### Chapter 16 Gene Expression: Control in Eukaryotes

**1.** See figure 16.5.

**3.** Since 5-azacytidine prevents methylation, the observed increase in transcription suggests that the presence of methyl groups inhibits transcription.

**5.** Genomic equivalence means that all of the cells of a multicellular eukaryotic organism are genetically identical. Yet, cells in different tissues and different regions of the organism are phenotypically different, expressing different suites of genes. Explaining differences in gene expression among cells that are genetically identical is a major question of eukaryotic genetics.

**7.** The three classes of segmentation genes in *Drosophila* are gap, pair-rule, and segment polarity. Mutations in gap genes leave gaps of missing segments. Mutations in pair-rule genes leave gaps of even or odd sets of segments. Mutations of segment polarity genes cause changes in all segments, generally the change in anterior or posterior portions of each.

**9.** In the development of the early *Drosophila* embryo, a syncitial blastoderm stage is achieved after thirteen cell divisions. The nuclei are near the surface of the embryo but not surrounded by cell membranes. Thereafter, membranes form, creating a cellular blastoderm.

**11.** Maternal-effect genes determine four regions of the developing embryo (major gene in parentheses): anterior (*bicoid*), posterior (*nanos*), dorso-ventral (*Toll*), and terminal (*torso*).

**13.** The helix-turn-helix motif (see box 16.1, fig. 1) consists of two alpha helices separated by a short turn within the protein, providing the structure to interact with DNA. Two other motifs are the zinc finger and the leucine zipper. Another motif, a combination of helix-turn-helix and leucine zipper, is shown in box 16.1, figure 4.

**15.** Amphibians have very large eggs, development is external to the female, a ready supply of zygotes is available, and they are easily manipulated experimentally.

**17.** Assuming that each cancer might be controlled by a single locus and assuming that breast cancer appears only in women and prostate cancer appears only in men (sex limited), pancreatic and prostate cancer are probably controlled by autosomal recessive genes; colon cancer is probably controlled by a dominant gene (autosomal or sex linked); and breast cancer by a recessive gene, either autosomal or sex linked.

**19.** The protein product of the retinoblastoma gene, p105, binds with oncogene proteins. Thus, the protein may somehow suppress transformation; when bound by oncogene proteins, p105 may be rendered ineffective. Hence, p105 seems to act to suppress transformation and thus the gene is called an anti-oncogene.

**21.** Animal viruses can have DNA or RNA, either single- or double-stranded. They can be enveloped or nonenveloped. They can have simple or complex protein coats.

**23.** The following are translation mechanisms: normal translation; read-through translation; and splice and then translation.

**25.** The v and c refer to viral and cellular, respectively. A proto-oncogene is a cellular oncogene within a nontransformed cell.

**27.** The v-*src* gene has no introns and the virus can function without the gene.

**29.** We see one band that is common to both cell lines; this band must represent the normal oncogene. The fact that this band is present in both lines indicates that the insertion of a virus has occurred in only one of the two copies of the gene present in the clone 1 cell. If it had inserted within both genes, we should not have seen the normal band. We see a larger fragment in clone 1, indicating that the DNA of the virus does not contain a site for the restriction enzyme used and that the virus has inserted within the restriction sites that define the band, lengthening the region probed. Alternatively, the virus could contain a restriction site and has still inserted in such a way as to lengthen the band probed by inserting between the sequence probed and the original restriction site.

**31.** The components of an immunoglobulin light chain are V, J, and C regions; the components of an immunoglobulin heavy chain are V, D, J, and C regions.

**33.** The V-J joining recognition signal is a heptamer and nonamer separated by twenty-three and twelve base pairs; see figure 16.35.

**35.** A T-cell receptor is an immunoglobulinlike molecule located on the surfaces of T cells, enabling them to identify infected host cells.

**37.** One explanation is a defect in the maturation process of B cells. Another explanation is a defect in the process of V(D)J joining, which could involve five or more genes.

**39.** The simplest interpretation of these results is that something is different in the organization of antibody genes in embryonic cells and in B lymphocytes. If the genes were in the same place in both cases, we should have seen identical patterns for the two types of cells. The probe recognizes both variable and constant regions of the antibody gene, since it is made from the mature mRNA. In the B lymphocyte, the variable and constant regions are adjacent, but in the embryonic cells, there is some extra DNA between these two genes. This result led to the notion that variable genes are rearranged during the development of the immune system.

*Critical Thinking Question:*

1. Adenovirus attacks normal cells by binding to p53; it also attacks cancerous cells that lack p53. If we remove the gene for the protein that binds p53, the *E1B* gene, then the modified adenovirus will not be able to attack normal cells but will be able to attack cancerous ones lacking p53. This modification of adenovirus as a potential tool in treating cancer was published in 1996.

### Chapter 17 Non-Mendelian Inheritance

1. Persistence of an environmentally induced trait into later generations is known as dauermodification and does not imply genetic control. After a suitable number of generations, the phenotype returns to normal, indicating an environmental rather than genetic response.
3. Maroon-like must affect the cytoplasm of the egg. The first cross is unusual and alerts us to maternal inheritance. A true Mendelian factor should produce one-half maroon-like males and females. The genotypes of the $F_1$ females are $ma\text{-}1^+/ma\text{-}1$ and $ma\text{-}1/ma\text{-}1$; half of the females should be of each genotype. If the wild-type allele produces wild-type cytoplasm for the progeny, regardless of the genotype of the progeny, any female that is $ma\text{-}1^+/ma\text{-}1$ will produce all wild-type progeny.
5. Although the genetic scheme predicts shell coiling perfectly, one could do experiments involving the injection of cytoplasm into eggs to test the viral hypothesis.
7. Female parent: *Dd.* Male parent: *d*–. (A): *dd.* Since selfing produces only sinistral snails, individual (A) must be homozygous for sinistral coiling, *dd.* The individual must get one *d* allele from each parent, so each parent must be at least heterozygous. Since (A) has dextral coiling, the mother must be heterozygous. The father could be either *D/d* or *d/d.*
9. By looking at different species, it is clear that few genes for oxidative phosphorylation are found in all mitochondrial genomes.
11. The rule of thumb is that suppressive petite mitochondria will dominate a cell, whereas neutral petite mitochondria will be lost in a competitive situation. Therefore

    segregational petite × segregational petite → segregational petites
    segregational petite × neutral petite → segregational petites
    segregational petite × suppressive petite → suppressive petites
    neutral petite × neutral petite → neutral petites
    neutral petite × suppressive petite → suppressive petites
    suppressive petite × suppressive petite → suppressive petites
13. In 0.02% of the offspring cells, the $mt^-$ allele of the streptomycin locus is inherited. In essence, these cells seem to have inherited chloroplast genes from the $mt^-$ parent. These cells thus provide us with a window on the possibility of having chloroplast genotypes from both parents viable within the same cell. Thus, interaction among other chloroplast genes can be looked for in this class of offspring (0.02% of total). If recombination occurs, map distances can be calculated by the usual methods, keeping in mind that we are taking data only from within this 0.02% of offspring.
15. Mitochondria and chloroplasts have prokaryotic affinities. They both have circular chromosomes, and their metabolism is affected by prokaryotic inhibitors (e.g., antibiotics). Certain prokaryotic messenger RNAs will hybridize with the organelle's DNA. There are numerous other aspects of biochemistry, morphology, and physiology that help to demonstrate affinities.
17. One way to determine that two loci are involved is to look at the proportion of offspring that lose mu particles after autogamy. In some strains, one-half the offspring will lose mu particles, indicating one locus was initially segregating (autogamy of $M_1m_1m_2m_2$ yields $M_1M_1m_2m_2$ or $m_1m_1m_2m_2$ in a 1:1 ratio). In other strains, one-fourth of the offspring will lose mu after autogamy, indicating that two unlinked loci were segregating (autogamy in $M_1m_1M_2m_2$ yields $M_1M_1M_2M_2$, $M_1M_1m_2m_2$, $m_1m_1M_2M_2$, or $m_1m_1m_2m_2$ in a 1:1:1:1 ratio).
19. Human mitochondrial DNA does not have introns. Finding an intron would suggest that the mitochondria had acquired a nuclear gene.
21. **a.** All type 1, 1.5 and 3.7 kilobases. **b.** All type 2, 2.5 and 6.0 kilobases. Recall that the chloroplast DNA from $mt^-$ cells does not appear in progeny.
23. Conjugation produces exconjugants with the same genotype, but which haploid micronucleus survives in each cell is random. Therefore, one-fourth of the time, the genotypes of the exconjugants are expected to be *KK,* one-half of the time *Kk,* and one-fourth of the time *kk.* The sensitive exconjugant remains sensitive, regardless of which of the genotypes is present. Autogamy does not affect the genotype of homozygous exconjugants, but it does affect the heterozygote. Among the killer exconjugants we expect:

| Killer Exconjugants | Autogamous Products | Phenotypic Ratio |
|---|---|---|
| 1/4 *KK* | all *KK* | 1/4 killer |
| 1/2 *Kk* | 1/2 *KK* | 1/4 killer |
| | 1/2 *kk* | 1/4 sensitive (Kappa are lost) |
| 1/4 *kk* | all *kk* | 1/4 sensitive (Kappa are lost) |

25. Use the striped plant as the egg parent, and get pollen from a plant with the following genotype: *IjIj jj.* If the striped plant is *iojap* (*ijij JJ*), all progeny will be heterozygous for both genes and will also contain *iojap* cytoplasm. The $F_1$ plants will segregate green, striped, and white sections within the plant. If the original plant is *IjIj jj* (and thus *japonica*), all $F_1$ plants will have striped leaves.
27. **a.** 2 normal:2 petite **b.** 0 normal:4 petite **c.** 4 normal: 0 petite. A nuclear gene should segregate 2:2 for each allele. Cytoplasmic factors will produce four spores with identical cytoplasm.

*Critical Thinking Question:*

1. We believe there are two mechanisms to ensure the distribution of cellular organelles during cytokinesis: stochastic and ordered inheritance. Stochastic inheritance simply means that no real mechanism exists; rather, the cell depends on the large number of the organelles to ensure an even distribution during the dividing of the cell. Ordered inheritance requires the even distribution of organelles in small numbers. This can be accomplished by special structures that divide a large organelle (e.g., a single, large chloroplast), or by other mechanisms that insert part of the mitochondrial system into new buds in budding yeast.

### Chapter 18 Quantitative Inheritance

1. Three in two hundred is approximately 1 in $64 = 1/(4)^3$; therefore, three loci (see table 18.1). Each effective allele contributes about 1/2 pound over the 2-pound base (3-pound difference divided by six effective alleles: *AA BB CC* = 5 pounds, *aa bb cc* = 2 pounds).
3. Independent assortment: for example, *Aa Bb Cc Dd* parents can have *AA BB CC DD* offspring.
5. Individuals of intermediate color can produce both lighter and darker offspring by independent assortment. That is, *Aa Bb Cc Dd*

parents can produce *AA BB CC DD* and *aa bb cc dd* offspring. However, if each effective allele adds color, then individuals with the base color (white) who mate with each other (*aabbccdd*) presumably cannot have children with darker skin.

**7.** The marker stock is exposed to DDT over many generations, selecting for DDT resistance. It is then crossed with the wild-type and $F_1$ offspring are backcrossed to generate flies with various combinations of selected chromosomes. These flies are then tested for their DDT resistance. For example:

$P_1$ *Cy/Pm H/S Ce/M*(4) × + (wild-type)
$F_1$ (male) *Cy*/+ *H*/+ *Ce*/+ × (backcross) + (wild-type female)
$F_2$ *Cy*/+ +/+ +/+; +/+ H/+ +/+, etc.

or,

$F_1$ (male) *Cy*/+ *H*/+ *Ce*/+ × (backcross) *Cy/Pm H/S Ce/M(4)* (female)
$F_2$ *Cy/Pm H*/+ *Ce*/+; +*/Pm* +*/S* +*/M(4)*; etc.

**9.** 50 cm. The difference between the two heights is 20 cm. This difference must result from the presence of effective (uppercase) alleles. The tall plant has four effective alleles, so each effective allele contributes an average of 5 cm to the height of the plant. The heterozygote has two effective alleles; 2 × 5 = 10 cm above the base of 40 cm, or a total of 50 cm.

**11.** Eight. The frequency of individuals in the $F_2$ that resemble one parent is $1/4^n$, where $n$ = the number of genes involved. In this case, 100 of 6,200,000 were like one parent. 100/6,200,000 is approximately 1/64,000, which approximates $1/4^8$. So we probably have eight genes involved.

**13.** $r = 0.43$; $H_N = 0.43/0.50 = 0.86$ (narrow-sense heritability); environmental variance, a component of the total phenotypic variance appears in the denominator of the heritability equations (18.11 and 18.12). Thus, environmental factors that lower environmental variance (more uniform environments or environmental effects) increase heritability. And, factors that raise the environmental variance (less uniform environments or environmental effects) decrease heritability.

**15.** $H = (4 - 0.9)/(5 + 5) = 0.31$; H (high line) $= (4 - 3)/5 = 0.2$; H (low line) $= (3 - 0.9)/5 = 0.42$. Part of the difference may be due to the number of alleles available for selection in each direction and nonadditive factors.

**17.** First, "outstanding athletic ability" must be defined. It can be defined subjectively by accomplishment or more objectively with a physiological measure. Then genetic effects must be assessed through heritability analyses such as twin studies and correlations among relatives.

**19.** Uniformly good nutrition should increase the heritability of height by eliminating some of the environmental variance; it affects only the denominator in a heritability equation.

**21.** Thorax length: $V_D + V_I = (100 - [43 + 51]) = 6$; $H_N = 43/100 = 0.43$; $H_B = 49/100 = 0.49$. Eggs laid: $V_D + V_I = 44$; $H_N = 0.18$; $H_B = 0.62$.

**23.** 4.5 g.

$$H_N = \frac{\text{gain}}{\text{selection differential}}$$

Thus ($H_N$) (selection differential) = gain. Since $H_N = 0.5$ and selective differential = 9, (0.5)(9.0) = 4.5 g.

**25.** 176 pounds. To solve this problem, use the formula for realized heritability:

$$H = \frac{\text{gain}}{\text{selection differential}} = \frac{(Y_0 - \overline{Y})}{(Y_P - \overline{Y})}$$

$$0.4 = \frac{(Y_0 - \overline{Y})}{(Y_P - \overline{Y})} = \frac{Y_0 - \overline{Y}}{185 - 170}$$

$$Y_0 - \overline{Y} = (0.4)(15) = 6.0$$

$$6.0 = Y_0 - 170$$
$$Y_0 = 176 \text{ lbs}$$

*Critical Thinking Question:*

**1.** The simplest cause for retraction of a study is the failure of that study to be replicated by others. That would come about when the conclusion isn't generally supportable, a phenomenon that could have two major causes. First, the study might have been done well but the phenomenon was specific to that study, due possibly to small sample sizes or a "private mutation," an effect found in one family or a small group of individuals but not a general phenomenon. Second, there could have been an inadvertent error in the study. For example, a size difference in a small region of the brain of homosexual and heterosexual males was reported but has not been verified. One investigator, who is following up the study, suggested that the difference could be an effect of differences in the way the brains were preserved and thus represent no real (biological) effect.

### Chapter 19 Population Genetics: The Hardy-Weinberg Equilibrium and Mating Systems

**1.** The frequencies of the three genotypes are $f(MM) = 41/100 = 0.41$; $f(MN) = 38/100 = 0.38$; $f(NN) = 21/100 = 0.21$. The frequency of *M*, *p*, is the frequency of *MM* homozygotes plus half the frequency of heterozygotes:

$$p = f(M) + (1/2)f(MN) = 0.41 + (1/2)(0.38) = 0.41 + 0.19 = 0.60$$
$$q = 1 - p = 1 - 0.60 = 0.40$$

Alternatively

$$p = f(M) = \frac{2(\#MM) + \#MN}{2 \times \text{total}} = \frac{2(41) + 38}{200}$$
$$= \frac{120}{200} = 0.60$$
$$q = 1 - p = 0.40$$

We do the following chi-square test:

| | *MM* | *MN* | *NN* | **Total** |
|---|---|---|---|---|
| Observed | 41 | 38 | 21 | 100 |
| Expected | $p^2 \times 100$ | $2pq \times 100$ | $q^2 \times 100$ | |
| | 36 | 48 | 16 | 100 |
| Chi-square | 0.694 | 2.083 | 1.563 | 4.340 |

The critical chi-square (0.05, one degree of freedom) = 3.841. We thus reject the null hypothesis that this population is in Hardy-Weinberg proportions.

**3.** Here we must assume Hardy-Weinberg equilibrium because of dominance. The $f(tt) = 65/215 = 0.302$. Thus $q = f(t) = \sqrt{f(tt)} = \sqrt{0.302} = 0.55$; and $p = f(T) = 1 - 0.55 = 0.45$. Since there are zero degrees of freedom (number of phenotypes − number of alleles = 2 − 2 = 0), we cannot do a chi-square test to determine if the population is mating at random (in Hardy-Weinberg proportions).

**5.** **a.** Most human populations should be in Hardy-Weinberg proportions at the taster locus regardless of what the allelic frequencies are. 3:1 is a family ratio when the parents are heterozygotes or a population ratio when $p = q = 0.5$, given dominance.

**b.** With no other information, it is probably safest to assume $p = q = 0.5$. At equal frequencies of alleles, the dominant phenotype (tasting) occurs in 75% of people ($p^2 + 2pq$).

**7.** $p = 0.99, q = 0.01$. If the population is in equilibrium, there should be $p^2$ of $AA$ + $2pq$ of $Aa$ + $q^2$ of $aa$ individuals. Since 1/10,000 shows the recessive trait, this is $q^2$. Therefore,

$$q = \sqrt{1/10{,}000} = \sqrt{0.0001} = 0.01$$

Since $p + q = 1, p = 1 - 0.01 = 0.99$.

**9.** The expected frequency of brown-eyed individuals will depend on the allelic frequencies of the original population. If we assume that mating is random with respect to eye color, and there is no selection, allelic frequencies will not change with time. We can, for example, calculate the frequencies of brown-eyed individuals for two populations at equilibrium. Let $p$ = frequency of the brown-eye allele and $q$ = frequency of the blue-eye allele.

| ***Population*** | ***p*** | ***q*** | ***Frequency of brown ($p^2 + 2pq$)*** |
|---|---|---|---|
| 1. | 0.7 | 0.3 | 0.91 |
| 2. | 0.5 | 0.5 | 0.75 |

We see that the original premise will be met only if the alleles are equally frequent.

**11.** 0.187 *MM,* 0.491 *MN,* 0.321 *NN.* The next generation will achieve equilibrium and there will be $(0.43)^2$ *MM* + $2(0.43)(0.57)$*MN* + $(0.57)^2$*MN.*

**13.** $f(A) = 0.792; f(B) = 0.208$. The easiest way to calculate frequencies is to do it empirically. We have three hundred people, so we have six hundred alleles.

$$f(A) = \frac{2 \times 200(AA) + 75(AB)}{600} = \frac{475}{600} = 0.792$$

$$f(B) = \frac{2 \times 25(BB) + 75(AB)}{600} = \frac{125}{600} = 0.208$$

**15.** **a.** Hardy-Weinberg proportions are achieved in one generation of random mating (multiple-allelic extension) if the locus is autosomal. If the locus is autosomal but there are different frequencies in the two sexes, then Hardy-Weinberg proportions are achieved in two generations. If the locus is sex linked, with different initial frequencies in the two sexes, then approach to equilibrium is gradual.

**b.** If the loci are not in equilibrium to begin with (linkage disequilibrium), then equilibrium is achieved asymptotically.

**17.** 0.63 type A, 0.08 type B, 0.28 type AB, 0.01 type O. Since the population is in equilibrium, the genotypic frequencies can be calculated as $(p + q + r)^2 = p^2 + 2pq + q^2 + 2pr + 2qr + r^2$. Let $p = 0.7$, $q = 0.2$, and $r = 0.1$. Blood types will be represented by the following:

| | | | |
|---|---|---|---|
| A: $p^2 + 2pr$ | B: $q^2 + 2qr$ | AB: $2pq$ | O: $r^2$ |
| $= 0.49 + 0.14$ | $= 0.04 + 0.04$ | $= 0.28$ | $= 0.01$ |

**19.** Inbreeding is disadvantageous when it causes recessive deleterious alleles to become homozygous. This occurs in normally outbred populations of diploids that have built up these harmful alleles. In species that normally inbreed, these deleterious alleles are probably no longer present; they were either removed by selection long ago or cannot build up in the population because of the regular pattern of inbreeding.

**21.** There are four paths passing through A, the only common ancestor ($F_A = 0.01$). All have ABCI as one side of the path. The second legs of the four other paths are ADEHI, ADGHI, AFGHI, and AFEHI. (A path such as IHEDAFGHI is invalid, passing through H twice.) Since each path has six ancestors, the inbreeding coefficient is $F_1 = 4(1/2)^6(1.01) = 0.063$.

**23.** Using the formula $F = (2pq - H)/2pq$, we calculate that $2pq = 0.48$, and $H = 38/100 = 0.38$. Therefore, $F = (0.48 - 0.38)/0.48 = 0.208$.

**25.** 0.375. $F = (2pq - H)/2pq = \dfrac{(0.32 - 0.20)}{0.32} = \dfrac{0.12}{0.32} = 0.375$

*Critical Thinking Question:*

**1.** Let us use the superscripts $m$ and $p$ for male and female, respectively. Then, we can construct the following Punnett square creating the next generation:

| | **Males,** $A; f(A) = p^m$ | **Males,** $a; f(a) = q^m$ |
|---|---|---|
| Females, $A; f(A) = p^f$ | $f(AA) = p^m p^f$ | $f(Aa) = p^f q^m$ |
| Females, $a; f(a) = q^f$ | $f(Aa) = p^m q^f$ | $f(aa) = q^m q^f$ |

Since the distribution of offspring in the table is independent of sex, it is the same in both sexes. The frequency of the $A$ allele, $p$ (in both sexes), will be the sum of the frequencies of the homozygotes and half the heterozygotes, or:

$$p = p^m p^f + (1/2)(p^m q^f + p^f q^m)$$

We then substitute $(1 - p)$ for all $q$'s:

$$p = p^m p^f + (1/2)p^m(1 - p^f) + (1/2)p^f(1 - p^m)$$

which simplifies to:

$$p = (1/2)(p^m + p^f)$$

In other words, after one generation of random mating, the allelic frequencies in each sex are the averages for both sexes. Now the frequency is the same in both sexes, and a second generation of random mating will achieve Hardy-Weinberg proportions. The population is not in equilibrium after one generation because the proportion of genotypes is not $p^2$, $2pq$, and $q^2$ if $p^m$ does not equal $p^f$. In other words, $p^m p^f$ does not equal $p^2$.

### Chapter 20 Population Genetics: Processes That Change Allelic Frequencies

**1.** **a.** The equilibrium frequency of $a$ is $\hat{q} = \dfrac{\mu}{(\mu + \nu)}$. Therefore

$$\hat{q} = (6 \times 10^{-5})/(6 \times 10^{-5} + 7 \times 10^{-7}) = 0.00006/0.0000607 = 0.988$$

**b.** If $q = 0.90$ at generation $n$, then

$$q_{n+1} = q_n + \mu p_n - \nu q_n = 0.90 + (6 \times 10^{-5})(0.10) - (7 \times 10^{-7})(0.90) = 0.9000054$$

**3.** 0.714.

$$\hat{q} = \frac{\mu}{(\mu + \nu)} = \frac{5 \times 10^{-5}}{5 \times 10^{-5} + 2 \times 10^{-5}} = \frac{5}{7} = 0.714$$

**5.** We use equation 20.12 to calculate the migration rate, *m:*

$$m = (q_C - q_N)/(q_M - q_N)$$

In this case, $q_C = 0.45$, $q_N = 0.62$, and $q_M = 0.03$. Thus

$$m = (0.45 - 0.62)/(0.03 - 0.62) = (-0.17)/(-0.59) = 0.288$$

**7.** Fast: 0.56; slow: 0.44. With 900 butterflies we have 1,800 alleles. 0.6 (1,800) = 1,080 fast alleles, and 0.4 (1,800) = 720 slow alleles. In the migrant population, (0.8)(180) = 144 slow and (0.2)(180) = 36 fast alleles. Therefore, the frequency of the fast allele is (1,080 + 36)/1,980 = 0.56. The frequency of the slow allele is 1 − f(fast allele) = 1 − 0.56 = 0.44.

**9.** 0.47. We again use equation 20.12:

$$m = \frac{q_C - q_N}{q_M - q_N}$$

where $m = 0.1$, $q_C = 0.45$, and $q_M = 0.25$.

$$0.1 = \frac{0.45 - q_N}{0.25 - q_N}$$

$$0.1(0.25 - q_N) = 0.45 - q_N$$

$$0.025 - 0.1q_N = 0.45 - q_N$$

$$0.9q_N = 0.425$$

$$q_N = \frac{0.425}{0.9} = 0.47$$

**11.** In stabilizing selection, extremes of a distribution are selected against. In directional selection, one extreme is favored over the other. In disruptive selection, both extremes are favored over the middle of the distribution.

**13.** Heterozygote disadvantage:

| | ***AA*** | ***Aa*** | ***aa*** | **Total** |
|---|---|---|---|---|
| Before selection | $p^2$ | $2pq$ | $q^2$ | 1 |
| Fitnesses ($W$) | 1 | $1 - s$ | 1 | |
| Frequencies after selection | $p^2/\overline{W}$ | $2pq(1 - s)/\overline{W}$ | $q^2/\overline{W}$ | $\overline{W} = 1 - 2pqs$ |

Then

$$q_{n+1} = (pq[1 - s] + q^2)/\overline{W}$$

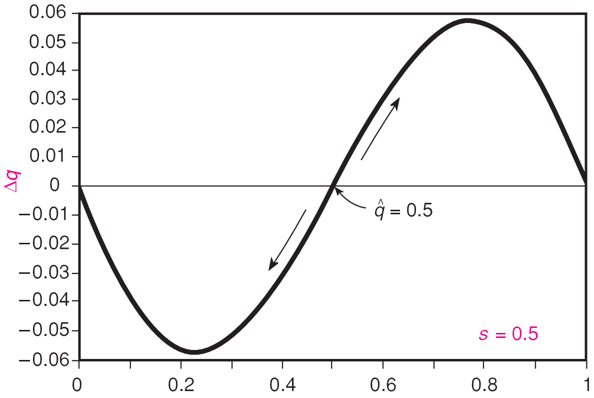
$$\Delta q = q_{n+1} - q = (pq[1 - s] + q^2)/\overline{W} - q\overline{W}/\overline{W}$$

$$= (pq[1 - s] + q^2 - q[1 - 2pqs])/\overline{W}$$

which simplifies to

$$\Delta q = spq(2q - 1)/\overline{W}$$

At $\Delta q = 0$, $\hat{q} = 0, 1$, or 0.5 ($2q - 1 = 0$, therefore $\hat{q}$ =0.5). The equilibrium points of zero and one are stable—if perturbed slightly (less than 0.5), the population will return to these values. The value $\hat{q}$ =0.5 is, however, unstable—if perturbed, it will continue away from the equilibrium point. This can be seen by either substituting into or graphing the $\Delta q$ equation.



**15.**

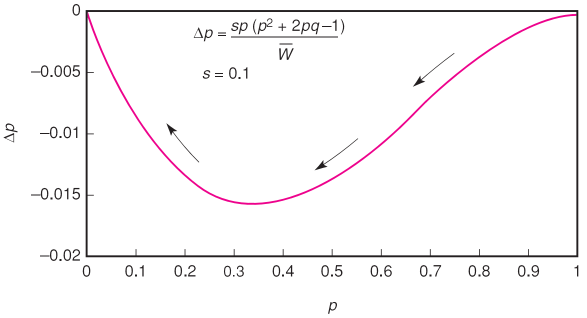
| | ***AA*** | ***Aa*** | ***aa*** | **Total** |
|---|---|---|---|---|
| Before selection | $p^2$ | $2pq$ | $q^2$ | 1 |
| Fitnesses ($W$) | $1 - s$ | $1 - s$ | 1 | $\overline{W} = 1 - p^2s - 2pqs$ |
| Frequencies after selection | $p^2(1 - s)/\overline{W}$ | $2pq(1 - s)/\overline{W}$ | $q^2/\overline{W}$ | 1 |

$$p_{n+1} = (p^2[1 - s] + pq[1 - s])/\overline{W}$$

$$\Delta p = p_{n+1} - p = (p^2[1 - s] + pq[1 - s])/\overline{W} - p\overline{W}/\overline{W}$$

$$= sp(p^2 + 2pq - 1)/\overline{W} = sp(- p^2 + 2p - 1)/\overline{W}$$

And $\hat{p} = 0$ or 1 (from the root of the quadratic). Only zero is stable.



**17.** Since only heterozygotes survive, $\hat{q} = 0.5$. This can also be derived from equation 20.24:

$$\hat{q} = s_1/(s_1 + s_2)$$

$$\text{If } s_1 = s_2 = 1, \text{ then } \hat{q} = 1/2$$

**19.** The *ST* inversion seems to do best at lower elevations and the *AR* at higher elevations. *CH* (and others) do not appear to be affected by altitude. To test this hypothesis, we would grow caged populations of flies with different initial frequencies of the various inversions at different simulated elevations, simulated by temperature, pressure, oxygen content, or other. We predict that, regardless of initial conditions, they would eventually equilibrate at the values in the table for the given parameter of the altitude (temperature, pressure, oxygen content, or other) that is acting as a selective agent. We would thus identify the selective agent. Since the inversions isolate various allelic combinations, our next step (a potentially long-term step) would be to determine which loci the selective agent is acting on.

**21.** 0.575. We can use the formula in equation 20.15, which is a simplified form of the sum of homozygotes + one-half the proportion of heterozygotes of the *b* allele.

$$q_{n+1} = \frac{q - sq^2}{1 - sq^2}$$

Since the relative fitness, $W = 0.4, s = 1 - W = 0.6$.

$$q_{n+1} = \frac{0.7 - (0.6)(0.49)}{1 - (0.6)(0.49)} = \frac{0.406}{0.706} = 0.575$$

**23.** 0.33, 0.25. Since the fitness is zero, $s = 1$, and we can use equation 20.17: $q_1 = q_0/(q_0 + 1) = 0.5/1.5 = 0.33$. For the second generation, we substitute the first generation numbers: $q_2 = 0.33/1.33 = 0.248$.

*Critical Thinking Question:*

**1.** There could be several reasons why these systems are in existence. First, they could be in selection-mutation equilibrium. However, that would not account for the high frequencies of both alleles in human populations in the Rh system. Second, the polymorphism could be relatively new, somehow maintaining both alleles as the human population increased in recent times with natural selection not having enough time to eliminate one of the alleles. Third, although the Rh blood system could follow the heterozygous disadvantage model, selection could also be acting in other ways that might maintain the polymorphism. That is, aside from Rh incompatibility eliminating heterozygotes, other genotypic combinations could be favored under other circumstances. Finally, although one or the other allele is being eliminated in any one population due to heterozygous disadvantage, the constant mixing of human populations could be reintroducing the rarer allele.

### Chapter 21 Evolution and Speciation

**1.** Neo-Darwinism is the application of population genetics to Darwinian evolution. Darwinian evolution works as natural selection favors the most fit organisms in competition among the overproduced young of any species.

**3.** Each process lets reproductive isolating mechanisms evolve while some barrier to breeding arises (see fig. 21.3).

**5.** *Constraint* refers to the limitations on changes that can take place. Some changes result in nonfunctional proteins and enzymes and thus cannot be in a successful lineage. For example, many base changes that lead to new amino acids in enzyme active sites disrupt enzymatic activity. If these mutations take place, they are eliminated by natural selection.

**7.** Yes. Two distinct species should not yield fertile progeny. We see a great reduction in the numbers of offspring from hybrids, indicating that hybrid inviability is one isolating mechanism operating.

**9.** Punctuated equilibrium proposes that species remain unchanged for long periods of time and that major changes occur only periodically. If species A existed for ten million years and suddenly (geologically speaking) changed dramatically to species X and Y, there would be few fossils because of the relatively short time in which intermediate forms were present. Another argument is, barring an incredibly detailed and complete sequence (of which there are almost none), there will always be gaps in the fossil record.

**11.** Genetic variability can be maintained by heterozygote advantage (e.g., sickle-cell anemia in people); frequency-dependent selection (rare-male mating advantage in *Drosophila*); transient polymorphism (industrial melanism in moths during an increase or decrease in industrialization); life-stage selection, which often happens when comparing egg, larval, pupal, and adult mortalities in *Drosophila;* differential selection in heterogeneous environments, common in some land snails; and neutrality.

**13.** Presumably, in mammals, selection could involve types of substrates acted upon by electrophoretic variants; functioning at different pHs and ionic strengths in various cellular compartments; resistance to enzyme inhibitors; interaction with other proteins and membrane components; and others.

**15.** Use the formula $K = -\ln(1 - d/n)$, in which $d$ is the number of amino acid differences and $n$ is the total number of amino acid sites. Thus:

human being–dog: $K = -\ln(1 - 1/8) = 0.133$

human being–chicken: $K = -\ln(1 - 3/8) = 0.470$

and dog–chicken: $K = -\ln(1 - 3/8) = 0.470$

These values place human beings and dogs very close and both equally far from chickens, which is consistent with the known evolutionary relationships.

**17.** 0.33 *AA,* 0.49 *Aa,* 0.18 *aa.* The mean fitness of the population after selection is $0.5 + 2pq = 0.98$ (table 21.2). The new frequency of a genotype is its original frequency times its fitness, all divided by the mean fitness of the population. Or:

$$f(AA) = \frac{(0.36)(1.5 - 0.6)}{0.98} = 0.33$$

$$f(Aa) = 0.48/0.98 = 0.49$$

$$f(aa) = (0.16)(1.5 - 0.4)/0.98 = 0.18$$

**19.** 0.023. $K = -\ln(1 - p)$

$K = -\ln(1 - [23/1{,}000]) = -\ln(0.0977) = 0.023$

**21.** The third. For many amino acids, the third position can be any of the four bases. In addition, wobble allows for some variation here as well. Changing the first or second base almost always produces a new amino acid.

**23.** Eight first cousins, on average, carry the complete genome of an individual. Therefore, from an evolutionary point of view, an individual and his or her eight cousins include the same alleles.

*Critical Thinking Question:*

**1.** Given that the gene for peppering is recessive and that one year equals one generation, the moths should be following a selection model in which natural selection acts against the recessive homozygote. In that case, we have already developed a selection model for this in chapter 20. Equation 20.15 relates the new allelic frequency to the old, given a particular selection coefficient:

$$q_{n+1} = q(1 - sq)/(1 - sq^2)$$

We can solve this equation for the selection coefficient:

$$s = (q - q_{n+1})/(q^2 - q^2q_{n+1})$$

Here, $q = 0.6$ and $q_{n+1} = 0.5$. When we solve the equation, we get: $s = 0.56$, or $W = 1 - s = 0.44$, which is the fitness of the peppered moth.

# APPENDIX B

## *Suggestions for Further Reading*

### Chapter 1 Introduction

Crow, J. 1993. N.I. Vavilov, martyr to genetic truth. *Genetics* 134:1–4.

Hull, D. 1996. A revolutionary philosopher of science. *Nature* 382:203–4.

Kuhn, T. 1962. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Maddox, J. 1999. The Unexpected Science to Come. *Scientific American,* December, 62–67.

Moore, J. 1993. *Science as a Way of Knowing.* Cambridge, Mass.: Harvard University Press.

Popper, K. 1962. *Conjectures and Refutations: The Growth of Scientific Knowledge.* New York: Basic Books.

Ruse, M., ed. 1989. *Philosophy of Biology.* New York: Macmillan.

Soyfer, V. 1994. *Lysenko and the Tragedy of Soviet Science.* New Brunswick, N.J.: Rutgers University Press.

Sterelny, K., and P. Griffiths. 1999. *Sex and Death: An Introduction to Philosophy of Biology.* Chicago: University of Chicago Press.

Stubbe, H. 1972. *History of Genetics.* Cambridge, Mass.: MIT Press.

Sturtevant, A. 1965. *A History of Genetics.* New York: Harper & Row.

### Chapter 2 Mendel's Principles

Beadle, G., and E. Tatum. 1941. Genetic control of biochemical reactions in *Neurospora. Proceedings of the National Academy of Sciences, USA* 27:499–506.

Bearn, A. 1994. Archibald Edward Garrod, the reluctant geneticist. *Genetics* 137:1–4.

Brink, R., ed. 1966. *Heritage from Mendel.* Madison: University of Wisconsin Press.

Corcos, A., and F. Monaghan. 1985. Role of de Vries in the recovery of Mendel's work. I: Was de Vries an independent discoverer of Mendel? *Journal of Heredity* 76:187–90.

———. 1987. Correns, an independent discoverer of Mendelism? I: An historical/critical note. *Journal of Heredity* 78:330.

———. 1993. *Gregor Mendel's Experiments on Plant Hybrids.* New Brunswick, N.J.: Rutgers University Press.

Douglas, L., and E. Novitski. 1977. What chance did Mendel's experiments give him of noticing linkage? *Heredity* 38:253–57.

Fisher, R. 1936. Has Mendel's work been rediscovered? *Annals of Science* 1:115–37.

Garrod, A. 1909. *Inborn Errors of Metabolism.* London: Hodder and Stoughton.

Greenspan, R. 1997. *Fly Pushing: The Theory and Practice of* Drosophila *Genetics.* Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press.

Horowitz, N. 1990. George Wells Beadle (1903–89). *Genetics* 124:1–6.

———. 1991. Fifty years ago: The *Neurospora* revolution. *Genetics* 127:631–35.

———. 1996. The sixtieth anniversary of biochemical genetics. *Genetics* 143:1–4.

Jackson, I. 1994. Molecular and developmental genetics of mouse coat color. *Annual Review of Genetics* 28:189–217.

Keightley, P. 1996. A metabolic basis for dominance and recessivity. *Genetics* 143:621–25.

King, R., and W. Stansfield. 1996. *Dictionary of Genetics,* 5th ed. New York: Oxford University Press.

Levy, H. 1999. Phenylketonuria: Old disease, new approach to treatment. *Proceedings of the National Academy of Sciences, USA* 96:1811–13.

Martin, D., W. Proebsting, and P. Hedden. 1997. Mendel's dwarfing gene: cDNAs from the *Le* alleles and function of the expressed proteins. *Proceedings of the National Academy of Sciences, USA* 94:8907–11.

Matton, D., N. Nass, A. Clarke, and E. Newbigin. 1994. Self-incompatibility: How plants avoid illegitimate offspring. *Proceedings of the National Academy of Sciences, USA* 91:1992–97.

Nasrallah, J., et al. 1994. Signaling the arrest of pollen tube development in self-incompatible plants. *Science* 266:1505–8.

Neel, J. 1994. *Physician to the Gene Pool: Genetic Lessons and Other Stories.* New York: Wiley.

Orel, V. 1996. *Gregor Mendel. The First Geneticist.* New York: Oxford University Press.

Perkins, D. 1992. *Neurospora:* The organism behind the molecular revolution. *Genetics* 130:687–701.

Peters, J. 1959. *Classical Papers in Genetics.* Englewood Cliffs, N.J.: Prentice Hall.

Pilgrim, I. 1986. A solution to the too-good-to-be-true paradox and Gregor Mendel. *Journal of Heredity* 77:218–20.

Reaume, A., D. Knecht, and A. Chovnick. 1991. The rosy locus in *Drosophila melanogaster:* Xanthine dehydrogenase and eye pigments. *Genetics* 129:1099–109.

Rick, C. 1991. Tomato paste: A concentrated review of genetic highlights from the beginnings to the advent of molecular genetics. *Genetics* 128:1–5.

Stern, C., and E. Sherwood, eds. 1966. *The Origin of Genetics, A Mendel Source Book.* San Francisco: Freeman.

Stubbe, H. 1972. *History of Genetics.* Cambridge, Mass.: MIT Press.

Sturtevant, A. 1965. *A History of Genetics.* New York: Harper & Row.

Thompson, R., and H.-H. Kirch. 1992. The S locus of flowering plants: When self-rejection is self-interest. *Trends in Genetics* 8:381–87.

Weiling, F. 1971. Mendel's "too good" data in *Pisum* experiments. *Folia Mendeliana* 6:75–77.

———. 1986. What about Fisher's statement of the "too good" data of J. G. Mendel's *Pisum* paper? *Journal of Heredity* 77:281–83.

Wellner, D., and A. Meister. 1981. A survey of inborn errors of amino acid metabolism and transport in man. *Annual Review of Biochemistry* 50:911–68.

Woolf, C., and J. Swafford. 1988. Evidence for eumelanin- and pheomelanin-producing genotypes in the Arabian horse. *Journal of Heredity* 79:100–106.

Yamamoto, F., et al. 1990. Molecular genetic basis of the histo-blood group ABO system. *Nature* 345:229–33.

Yoshida, K., et al. 1995. Cause of blue petal colour. *Nature* 373:291.

### Chapter 3 Mitosis and Meiosis

Barton, N., and L. Goldstein. 1996. Going mobile: Microtubule motors and chromosome segregation. *Proceedings of the National Academy of Sciences, USA* 93:1735–42.

Carpenter, A. 1994. Chiasma function. *Cell* 77:959–62.

Dawe, R. 1998. Meiotic chromosome organization and segregation in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 49:371–95.

Dobie, K., et al. 1999. Centromere proteins and chromosome inheritance: A complex affair. *Current Opinion in Genetics & Development* 9:206–17.

Glover, D., C. Gonzalez, and J. Raff. 1993. The centrosome. *Scientific American,* June, 62–68.

Grancell, A., and P. Sorger. 1998. Chromosome movement: kinetochores motor along. *Current Biology* 8:R382–85.

Heneen, W. 1982. The centromeric region in the scanning electron microscope. *Hereditas* 97:311–14.

Hirokawa, N., Y. Noda, and Y. Okada. 1998. Kinesin and dynein superfamily proteins in organelle transport and cell division. *Current Opinion in Cell Biology* 10:60–73.

Holm, C. 1994. Coming undone: How to untangle a chromosome. *Cell* 77:955–57.

Hoyt, M., and J. Geiser. 1996. Genetic analysis of the mitotic spindle. *Annual Review of Genetics* 30:7–33.

Hyams, J., and C. Lloyd, eds. 1994. *Microtubules.* New York: Wiley-Liss.

Kellogg, D., M. Moritz, and B. Alberts. 1994. The centrosome and cellular organization. *Annual Review of Biochemistry* 63:639–74.

Marc, J. 1997. Microtubule-organizing centres in plants. *Trends in Plant Science* 2:223–30.

Nasmyth, K. 1999. Separating sister chromatids. *Trends in Biochemical Sciences* 24:98–104.

Nicklas, R. 1997. How cells get the right chromosomes. *Science* 275:632–37.

Nurse, P. 2000. A long twentieth century of the cell cycle and beyond. *Cell* 100:71–78.

Orr-Weaver, T. 1999. The ties that bind: Localization of the sister-chromatid cohesin complex on yeast chromosomes. *Cell* 99:1–4.

Page, A., and P. Hieter. 1999. The anaphase-promoting complex: New subunits and regulators. *Annual Review of Biochemistry* 68:583–609.

Pigozzi, M., and A. Solari. 1999. Recombination nodule mapping and chiasma distribution in spermatocytes of the pigeon, *Columba livia. Genome* 42:308–14.

Rhoades, M. 1950. Meiosis in maize. *Journal of Heredity* 41:59–67.

Schubert, I., and J. Oud. 1997. There is an upper limit of chromosome size for normal development of an organism. *Cell* 88:515–20.

Skibbens, R., and P. Hieter. 1998. Kinetochores and the checkpoint mechanism that monitors for defects in the chromosome segregation machinery. *Annual Review of Genetics* 32:307–37.

Sutton, W. 1903. The chromosomes in heredity. *Biological Bulletin* 4:231–51.

Theriot, J., and L. Satterwhite. 1997. New wrinkles in cytokinesis. *Nature* 385:388–89.

von Wettstein, D., S. Rasmussen, and P. Holm. 1984. The synaptonemal complex in genetic segregation. *Annual Review of Genetics* 18:331–413.

Waters, J., and E. Salmon. 1997. Pathways of spindle assembly. *Current Opinion in Cell Biology* 9:37–43.

Whittaker, R. 1969. New concepts of kingdoms of organisms. *Science* 163:150–60.

Zickler, D., and N. Kleckner. 1998. The leptotene-zygotene transition of meiosis. *Annual Review of Genetics* 32:619–97.

———. 1999. Meiotic chromosomes: Integrating structure and function. *Annual Review of Genetics* 33:603–754.

Zimmerman, W., C. Sparks, and S. Doxsey. 1999. Amorphous no longer: The centrosome comes into focus. *Current Opinion in Cell Biology* 11:122–28.

## Chapter 4 Probability and Statistics

Berger, J., and D. Berry. 1988. Statistical analysis and the illusion of objectivity. *American Scientist* 76:159–65.

Dixon, W., and F. Massey, Jr. 1983. *Introduction to Statistical Analysis,* 4th ed. New York: McGraw-Hill.

Howson, C., and P. Urbach. 1991. Bayesian reasoning in science. *Nature* 350:371–74.

Jefferys, W., and J. Berger. 1992. Ockham's razor and Bayesian analysis. *American Scientist* 80:64–72.

Maynard Smith, J. 1968. *Mathematical Ideas in Biology.* New York: Cambridge University Press.

Moore, D. 1997. *Statistics: Concepts and Controversies,* 4th ed. San Francisco: Freeman.

Ross, S. 1994. *A First Course in Probability,* 4th ed. New York: Macmillan.

Siegel, S., and N. Castellan, Jr. 1988. *Nonparametric Statistics for the Behavioral Sciences,* 2d ed. New York: McGraw-Hill.

Yoccoz, N. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin of the Ecological Society of America* 72:106–11.

## Chapter 5 Sex Determination, Sex Linkage, and Pedigree Analysis

Baker, B., M. Gorman, and I. Marin. 1994. Dosage compensation in *Drosophila. Annual Review of Genetics* 28:491–521.

Baker, B., et al. 1991. The doublesex locus of *Drosophila melanogaster* and its flanking regions: A cytogenetic analysis. *Genetics* 127:125–38.

Barr, M., and E. Bertram. 1949. A morphological distinction between neurones of the male and female, and the behavior of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature* 163:676–77.

Bashaw, G., and B. Baker. 1996. Dosage compensation and chromatin structure in *Drosophila. Current Opinion in Genetics and Development* 6:496–501.

Bridges, C. 1932. The genetics of sex in *Drosophila.* In *Sex and Internal Secretions, A Survey of Recent Research,* edited by E. Allen, pp. 53–93. Baltimore: Williams & Wilkins.

Brockdorff, N., et al. 1991. Conservation of position and exclusive expression of mouse *Xist* from the inactive X chromosome. *Nature* 351:329–31.

Burgoyne, P. 1982. Genetic homology and crossing over in the X and Y chromosomes of mammals. *Human Genetics* 61:85–90.

Carrel, L., et al. 1999. A first-generation X-inactivation profile of the human X chromosome. *Proceedings of the National Academy of Sciences, USA* 96:14440–44.

Cline, T. 1988. Evidence that *sisterless-a* and *sisterless-b* are two of several discrete "numerator elements" of the X/A sex determination signal in *Drosophila* that switch *Sxl* between two alternative stable expression states. *Genetics* 119:829–62.

Cline, T., and B. Meyer. 1996. Vive la difference: Males vs. females in flies vs. worms. *Annual Review of Genetics* 30:637–702.

Corcos, A. 1983. Pattern baldness: Its genetics revisited. *American Biology Teacher* 45:371–75.

Crow, J. 1994. Advantages of sexual reproduction. *Developmental Genetics* 15:205–13.

Eicher, E., and L. Washburn. 1986. Genetic control of primary sex determination in mice. *Annual Review of Genetics* 20:327–60.

Elbrecht, A., and R. Smith. 1992. Aromatase enzyme activity and sex determination in chickens. *Science* 255:467–70.

Farrow, M., and R. Juberg. 1969. Genetics and laws prohibiting marriage in the United States. *Journal of the American Medical Association* 209:534–38.

Gepner, J., and T. Hays. 1993. A fertility region on the Y chromosome of *Drosophila melanogaster* encodes a dynein microtubule motor. *Proceedings of the National Academy of Sciences, USA* 90:11132–36.

Goodfellow, P., and R. Lovell-Badge. 1993. *SRY* and sex determination in mammals. *Annual Review of Genetics* 27:71–92.

Gorman, M., and B. Baker. 1994. How flies make one equal two: Dosage compensation in *Drosophila. Trends in Genetics* 10:376–80.

Graham, J., and R. Winters. 1961. Familial hypophosphatemia: An inherited demand for increased vitamin D. *Annals of the NY Academy of Sciences* 91:667–73.

Haqq, C., et al. 1993. *SRY* recognizes conserved DNA sites in sex-specific promoters. *Proceedings of the National Academy of Sciences, USA* 90:1097–101.

———. 1994. Molecular basis of mammalian sexual determination: Activation of Müllerian inhibiting substance gene expression by *SRY. Science* 266:1494–1500.

Harley, V., et al. 1992. DNA binding activity of recombinant *SRY* from normal males and XY females. *Science* 255:453–56.

Heard, E., P. Clerc, and P. Avner. 1997. X-chromosome inactivation in mammals. *Annual Review of Genetics* 31:571–610.

Herzing, L., et al. 1997. *Xist* has properties of the X-chromosome inactivation centre. *Nature* 386:272–75.

Hines, P., and E. Culotta. 1998. The evolution of sex. *Science* 281:1979–2008. [Eight news items and articles on the evolution of sex.]

Hunter, R. 1995. *Sex Determination, Differentiation and Intersexuality in Placental Mammals.* Cambridge, Mass.: Cambridge University Press.

Kelley, R., et al. 1997. *Sex lethal* controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature* 387:195–99.

———. 1999. Epigenetic spreading of the *Drosophila* dosage compensation complex from *roX* RNA genes into flanking chromatin. *Cell* 98:513–22.

Kohler, R. 1994. *Lords of the Fly*: Drosophila *Genetics and the Experimental Life.* Chicago: University of Chicago Press.

Koopman, P., et al. 1991. Male development of chromosomally female mice transgenic for *Sry. Nature* 351:117–21.

Lahn, B., and D. Page. 1999. Four evolutionary strata on the human X chromosome. *Science* 286:964–67.

Lindsley, D., and G. Zimm. 1992. *The genome of* Drosophila melanogaster. San Diego: Academic Press.

Long, A., and R. Michod. 1995. Origin of sex for error repair. I:Sex, diploidy, and haploidy. *Theoretical Population Biology* 47:18–55.

Lyon, M. 1962. Sex chromatin and gene action in the mammalian X chromosome. *American Journal of Human Genetics* 14:135–48.

———. 1992. Some milestones in the history of X-chromosome inactivation. *Annual Review of Genetics* 26:17–28.

———. 1999. X-chromosome inactivation. *Current Biology* 9:R235–37.

McKusick, V.A. 1998. *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders,* 12th ed. Baltimore: Johns Hopkins University Press.

Miller, O. 1995. The fifties and the renaissance in human and mammalian cytogenetics. *Genetics* 139:489–94.

Morgan, T. 1910. Sex-limited inheritance in *Drosophila. Science* 32:120–22.

Nicoll, M., C. Akerib, and B. Meyer. 1997. X-chromosome-counting mechanisms that determine nematode sex. *Nature* 388:200–204.

Parkhurst, S., and P. Meneely. 1994. Sex determination and dosage compensation: Lessons from flies and worms. *Science* 264:924–32.

Rice, W. 1994. Degeneration of a nonrecombining chromosome. *Science* 263:230–32.

———. 1996. Evolution of the Y sex chromosome in animals. *BioScience* 46:331–42.

Scott, D., et al. 1995. Identification of a mouse male-specific transplantation antigen, H-Y. *Nature* 376:695–98.

Shapiro, L., et al. 1979. Noninactivation of an X-chromosome locus in man. *Science* 204:1224–26.

Smithies, O. 1995. Early days of gel electrophoresis. *Genetics* 139:1–4.

Vainio, S., et al. 1999. Female development in mammals is regulated by Wnt-4 signaling. *Nature* 397:405–9.

Walker, C., et al. 1991. The Barr body is a looped X chromosome formed by telomere association. *Proceedings of the National Academy of Sciences, USA* 88:6191–95.

Willard, H. 1996. X chromosome inactivation, *XIST,* and pursuit of the X-inactivation center. *Cell* 86:5–7.

Wright, W. 1988. Sex change in the Mollusca. *Trends in Ecology and Evolution* 3:137–40.

### Chapter 6 Linkage and Mapping in Eukaryotes

Beuler, E., et al. 1996. Mutation nomenclature: Nicknames, systematic names, and unique identifiers. *Human Mutation* 8:203–6.

Bowring, F., and D. Catcheside. 1999. Evidence for negative interference: Clustering of crossovers close to the *am* locus in *Neurospora crassa* among *am* recombinants. *Genetics* 152:965–69.

Bridges, C. 1935. Salivary chromosome maps. *Journal of Heredity* 26:60–64.

Carson, S., W. Henry, and T. Shows. 1985. Tissue factor gene localized to human chromosome 1 (1pter → 1p21). *Science* 229:991–93.

Creighton, H., and B. McClintock. 1931. A correlation of cytological and genetical crossing over in *Zea mays. Proceedings of the National Academy of Sciences, USA* 17:492–97.

Crow, J. 1988. A diamond anniversary: The first chromosome map. *Genetics* 118:1–3.

———. 1990. Mapping functions. *Genetics* 125:669–71.

Emery, A., and D. Rimoin, eds. 1990. *Principles and Practice of Medical Genetics,* 2d ed. New York: Churchill Livingstone.

Federoff, N. 1993. Barbara McClintock. *Evolución Biológica* 7:1–19.

Fincham, J. 1998. Fungal genetics—Past and present. *Journal of Genetics* 77:55–63.

Greenspan, R. 1997. *Fly Pushing; The Theory and Practice of* Drosophila *Genetics.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Hall, M., and P. Linder. 1993. *The Early Days of Yeast Genetics*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Hayles, J., and P. Nurse. 1992. Genetics of the fission yeast *Schizosaccharomyces pombe. Annual Review of Genetics* 26:373–402.

Kaback, D., et al. 1999. Chromosome size-dependent control of meiotic reciprocal recombination in *Saccharomyces cerevisiae:* The role of crossover interference. *Genetics* 152:1475–86.

Kohler, R. 1994. *Lords of the Fly:* Drosophila *Genetics and the Experimental Life.* Chicago: University of Chicago Press.

Lefevre, G., and W. Watkins. 1986. The question of the total gene number in *Drosophila melanogaster. Genetics* 113:869–95.

Lewis, E. 1995. Remembering Sturtevant. *Genetics* 141:1227–30.

Lindsley, D., and G. Zimm. 1992. *The genome of* Drosophila melanogaster. San Diego: Academic Press.

Littlefield, J. 1964. Selection of hybrids from matings of fibroblasts in vitro and their presumed recombinants. *Science* 145:709–10.

Lyon, M. 1990. L. C. Dunn and mouse genetic mapping. *Genetics* 125:231–36.

Manney, T., and M. Manney. 1993. Using yeast genetics to generate a research environment. *Genetics* 134:387–91.

McKusick, V. 1998. *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders,* 12th ed. Baltimore: Johns Hopkins University Press.

McPeek, M., and T. Speed. 1995. Modeling interference in genetic recombination. *Genetics* 139:1031–44.

Morgan, T. 1919. *The Physical Basis of Heredity.* Philadelphia: Lippincott.

Morton, N. 1955. Sequential test for the detection of linkage. *American Journal of Human Genetics* 7:277–318.

———. 1995. LODs past and present. *Genetics* 140:7–12.

Munz, P. 1994. An analysis of interference in the fission yeast *Schizosaccharomyces pombe. Genetics* 137:701–7.

O'Brien, S. 1993. *Genetic Maps, 6th ed. Book 3: Lower Eukaryotes. Book 4: Nonhuman Vertebrates. Book 5: The Human Maps. Book 6: Plants.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Panthier, J., et al. 1990. Evidence for mitotic recombination in *Wei*/+ heterozygous mice. *Genetics* 125:175–82.

Preuss, D., S. Rhee, and R. Davis. 1994. Tetrad analysis possible in *Arabidopsis* with mutation of the QUARTET (QRT) genes. *Science* 264:1458–60.

Risch, N. 1992. Genetic linkage: Interpreting Lod scores. *Science* 255:803–4.

Roman, H. 1986. The early days of yeast genetics: A personal narrative. *Annual Review of Genetics* 20:1–12.

Ruddle, F., and R. Kucherlapati. 1974. Hybrid cells and human genes. *Scientific American,* July, 36–44.

Sorsa, V. 1988. *Chromosome Maps of* Drosophila. Two vols. Boca Raton, Fl.: CRC Press.

Sowers, A., ed. 1987. *Cell Fusion.* New York: Plenum Press.

Stahl, F., and R. Lande. 1995. Estimating interference and linkage map distance from two-factor tetrad data. *Genetics* 139:1449–54.

Stern, C. 1936. Somatic crossing over and segregation in *Drosophila melanogaster. Genetics* 21:625–730.

Sturtevant, A. 1913. The linear arrangement of six sex-linked factors in *Drosophila,* as shown by their mode of association. *Journal of Experimental Zoology* 14:43–59.

Sugawara, O., et al. 1990. Induction of cellular senescence in immortalized cells by human chromosome 1. *Science* 247:707–10.

Svinarich, D., and S. Krawetz. 1993. Gene assignment by quantitative hybridization analysis of somatic cell hybrids. *BioTechniques* 14:82–86.

Verma, R., and A. Babu. 1994. *Human Chromosomes. Manual of Basic Techniques.* Elmsford, N.Y.: Pergamon Press.

Zallen, D., and R. Burian. 1992. On the beginnings of somatic cell hybridization: Boris Ephrussi and chromosome transplantation. *Genetics* 132:1–8.

Zhao, H., and T. Speed. 1996. On genetic map functions. *Genetics* 142:1369–77.

### Chapter 7 Linkage and Mapping in Prokaryotes and Bacterial Viruses

American Association for the Advancement of Science. 1997. Frontiers in microbiology. *Science* 276:2. May, special issue.

Ankenbauer, R. 1997. Reassessing forty years of genetic doctrine: Retrotransfer and conjugation. *Genetics* 145:543–49.

Atlas, R. 1995. *Principles of Microbiology.* Dubuque, Iowa: Wm. C. Brown Publishers.

Bachmann, B. 1990. Linkage map of *Escherichia coli* K-12, 8th ed. *Microbiological Review* 54:130–97.

Berlyn, M., and S. Letovsky. 1992. COTRANS: A program for cotransduction analysis. *Genetics* 131:235–41.

Birge, E. 1994. *Bacterial and Bacteriophage Genetics,* 3d ed. New York: Springer-Verlag.

Blum, P., et al. 1999. Forward to the special issue on archaeal genetics. *Genetics* 152:1243. (Followed by six articles on the genetics of the archaea.)

Cairns, J., G. Stent, and J. Watson, eds. 1966. Phage and the origins of molecular biology. *Cold Spring Harbor Symposium on Quantitative Biology* 31.

Cavalli-Sforza, L. 1992. Forty years ago in *Genetics:* The unorthodox mating behavior of bacteria. *Genetics* 132:635–37.

Cerritelli, M., et al. 1997. Encapsidated conformation of bacteriophage T7 DNA. *Cell* 91:271–80.

Clewell, D. 1993. *Bacterial Conjugation.* New York: Plenum Press.

Eigen, M. 1993. Viral quasispecies. *Scientific American,* July, 42–49.

Garfield, E. 1990. A tribute to Joshua Lederberg: Highlights of a remarkable scientific career. *Current Contents, Life Sciences* 33:5–11.

Hoch, J. 1991. Genetic analysis in *Bacillus subtilis. Methods in Enzymology* 204:305–20.

Hoppert, M., and F. Mayer. 1999. Prokaryotes. *Scientific American,* November-December, 518–25.

Ippen-Ihler, K., and E. Minkley, Jr. 1986. The conjugation system of F, the fertility factor of *Escherichia coli. Annual Review of Genetics* 20:593–624.

Jacob, F., and E. Wollman. 1961. *Sexuality and the Genetics of Bacteria.* New York: Academic Press.

Lanka, E., and B. Wilkins. 1995. DNA processing reactions in bacterial conjugation. *Annual Review of Biochemistry* 64:141–69.

Lederberg, J. 1987. Genetic recombination in bacteria: A discovery account. *Annual Review of Genetics* 21:23–46.

———. 1989. Replica plating and indirect selection of bacterial mutants: Isolation of preadaptive mutants in bacteria by sib selection. *Genetics* 121:395–99.

Lederberg, J., et al. 1951. Recombination analysis of bacterial heredity. *Cold Spring Harbor Symposium on Quantitative Biology* 16:413–43.

Lloyd, R., and C. Buckman. 1995. Conjugational recombination in *Escherichia coli*: Genetic analysis of recombinant formation in Hfr X $F^-$crosses. *Genetics* 139:1123–48.

Low, K. 1991. Conjugational methods for mapping with Hfr and F-prime strains. *Methods in Enzymology* 204:43–62.

Mackenzie, C., A. Simmons, and S. Kaplan. 1999. Multiple chromosomes in bacteria: The Yin and Yang of *trp* gene localization in *Rhodobacter sphaeroides* 2.4.1. *Genetics* 153:525–38.

Marciano, D., M. Russel, and S. Simon. 1999. An aqueous channel for filamentous phage export. *Science* 284:1516–19.

Miller, J. 1992. *A Short Course in Bacterial Genetics.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Miller, R. 1998. Bacterial gene swapping in nature. *Scientific American,* January, 66–71.

Oldstone, M., and A. Levine. 2000. Virology in the next millennium. *Cell* 100:139–42.

Redfield, R., M. Schrag, and A. Dean. 1997. The evolution of bacterial transformation: Sex with poor relations. *Genetics* 146:27–38.

Stahl, F. 1989. The linkage map of phage T4. *Genetics* 123:245–48.

Stent, G., and R. Calendar. 1978. *Molecular Genetics, An Introductory Narrative,* 2d ed. San Francisco: Freeman.

Sternberg, N., and R. Maurer. 1991. Bacteriophage-mediated generalized transduction in *Escherichia coli* and *Salmonella typhimurium. Methods in Enzymology* 204:18–43.

Susman, M. 1995. The Cold Spring Harbor phage course (1945–70): A 50th anniversary remembrance. *Genetics* 139:1101–6.

Teleman, A., et al. 1998. Chromosome arrangement within a bacterium. *Current Biology* 8:1102–9.

Wikoff, W., and J. Johnson. 1999. Virus assembly: Imaging a molecular machine. *Current Biology* 9:R296–300.

Zinder, N. 1992. Forty years ago: The discovery of bacterial transduction. *Genetics* 132:291–94.

### Chapter 8 Cytogenetics

Antonarakis, S., et al. 1993. Mitotic errors in somatic cells cause trisomy 21 in about 4.5% of cases and are not associated with advanced maternal age. *Nature Genetics* 3:146–50.

Arnold, M. 1994. Natural hybridization and Louisiana irises. *BioScience* 44:141–47.

Ashley, C., Jr., and S. Warren. 1995. Trinucleotide repeat expansion and human disease. *Annual Review of Genetics* 29:703–28.

Ashworth, A., et al. 1991. X-chromosome inactivation may explain the difference in viability of XO humans and mice. *Nature* 351:406–8.

Barch, M., ed. 1997. *ACT Cytogenetics Laboratory Manual.* Philadelphia: Lippincott.

Bedell, M., N. Jenkins, and N. Copeland. 1996. Good genes in bad neighbourhoods. *Nature Genetics* 12:229–32.

Benavente, E., and J. Orellana. 1991. Chromosome differentiation and pairing behavior of polyploids: An assessment on preferential metaphase I associations in colchicine-induced autotetraploid hybrids within the genus *Secale. Genetics* 128:433–42.

Borgaonkar, D. 1997. *Chromosomal Variation in Man: A Catalogue of Chromosomal Variants and Anomalies,* 8th ed. New York: Alan R. Liss, Inc.

Brewster, T., and P. Gerald. 1978. Chromosome disorders associated with mental retardation. *Pediatric Annals* 7:82–89.

Chenuil, A., N. Galtier, and P. Berrebi. 1999. A test of the hypothesis of an autopolyploid vs. allopolyploid origin for a tetraploid lineage: Application to the genus *Barbus* (Cyprinidae). *Heredity* 82:373–80.

Cochran, D. 1983. Alternate-2 disjunction in the German cockroach. *Genetics* 104:215–17.

Coyne, J., et al. 1993. The fertility effects of pericentric inversions in *Drosophila melanogaster. Genetics* 134:487–96.

Cribiu, E., et al. 1989. Identification of chromosomes involved in a Robertsonian translocation in cattle. *Genetique, Selection, Evolution* 21:555–60.

De Vicente, M., and P. Arus. 1996. Tetrasomic inheritance of isozymes in sainfoin (*Onobrychis viciaefolia* Scop.). *Journal of Heredity* 87:54–62.

Dewald, G., et al. 1980. Origin of chi46,XX/46,XY chimerism in a human true hermaphrodite. *Science* 207:321–23.

Epstein, C. 1988. Mechanisms of the effects of aneuploidy in mammals. *Annual Review of Genetics* 22:51–75.

Futch, D. 1966. A study of speciation in South Pacific populations of *Drosophila ananassae. University of Texas Studies in Genetics* 6615:79–120.

Gallardo, M., et al. 1999. Discovery of tetraploidy in a mammal. *Nature* 401:341.

Jacobs, P., and T. Hassold. 1995. The origin of numerical chromosome abnormalities. *Advances in Genetics* 33:101–33.

Jones, G. 1994. Meiosis in autopolyploid *Crepis capillaris.* III: Comparison of triploids and tetraploids; evidence for non-independence of autonomous pairing sites. *Heredity* 73:215–19.

Karpechenko, G. 1928. Polyploid hybrids of *Raphanus sativus* L. X *Brassica oleracea* L. *Zeitschrift fuer Induktive Abstamungslehre und Vererbungslehre* 48:1–85.

Khush, G., et al. 1984. Primary trisomics of rice: Origin, morphology, cytology and use in linkage mapping. *Genetics* 107:141–63.

Koller, A., J. Heitman, and M. Hall. 1996. Regional bivalent-univalent pairing *versus* trivalent pairing of a trisomic chromosome in *Saccharomyces cerevisiae. Genetics* 144:957–66.

Leitch, I., and M. Bennett. 1997. Polyploidy in angiosperms. *Trends in Plant Science* 2:470–76.

Lejeune, J., M. Gautier, and R. Turpin. 1959. Etude des chromosomes somatiques de neuf enfants mongoliens. *Comptes Rendus de L'Academie des Sciences* (*Paris*) 248:1721–22.

Loidl, J. 1995. Meiotic chromosome pairing in triploid and tetraploid *Saccharomyces cerevisiae. Genetics* 139:1511–20.

Lukaszewski, A. 1995. Chromatid and chromosome-type breakage-fusion-bridge cycles in wheat (*Triticum aestivum* L.). *Genetics* 140:1069–85.

Martin, A., et al. 1999. A fertile amphiploid between diploid wheat (*Triticum tauschii*) and crested wheatgrass (*Agropyron cristatum*). *Genome* 42:519–24.

Orr, H. 1990. "Why polyploidy is rarer in animals than in plants" revisited. *American Naturalist* 136:759–70.

Page, S., and L. Shaffer. 1997. Nonhomologous Robertsonian translocations form predominantly during female meiosis. *Nature Genetics* 15:231–32.

Penrose, L. 1933. The relative effects of paternal and maternal age in mongolism. *Journal of Genetics* 27:219–24.

Pokholkova, G., et al. 1993. Observations on the induction of position effect variegation of euchromatic genes in *Drosophila melanogaster. Genetics* 133:231–42.

Richards, R., et al. 1992. Evidence of founder chromosomes in fragile X syndrome. *Nature Genetics* 1:257–60.

Sable, J., and S. Henikoff. 1996. Copy number and orientation determine the susceptibility of a gene to silencing by nearby heterochromatin in *Drosophila. Genetics* 142:447–58.

Sturtevant, A. 1925. The effects of unequal crossing over at the Bar locus in *Drosophila. Genetics* 10:117–47.

Talbert, P., and S. Henikoff. 2000. A reexamination of spreading of position-effect variegation in the *white-roughest* region of *Drosophila melanogaster. Genetics* 154:259–72.

Therman, E., and B. Susman. 1995. *Human Chromosomes,* 3d ed. New York: Springer-Verlag.

Trottier, Y., D. Devys, and J. Mandel. 1993. An expanding story (Fragile-X syndrome). *Current Biology* 3:783–86.

Warburton, D., J. Byrne, and N. Canki. 1991. *An Atlas of Prenatal Development in Conceptions with Chromosomal Anomalies.* New York: Oxford University Press.

Witkin, H., et al. 1976. Criminality in XYY and XXY men. *Science* 193:547–55.

Zheng, Y., R. Roseman, and W. Carlson. 1999. Time course study of the chromosome-type breakage-fusion-bridge cycle in maize. *Genetics* 153:1435–44.

### Chapter 9 Chemistry of the Gene

Abdurashidova, G., et al. 2000. Start sites of bi-directional DNA synthesis at the human lamin B2 origin. *Science* 287:2023–26.

Adams, D., et al. 1992. The role of topoisomerase IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. *Cell* 71:277–88.

Anderson, R., et al. 1996. Transmission dynamics and epidemiology of BSE in British cattle. *Nature* 382:779–88.

Arscott, P., et al. 1989. Scanning tunnelling microscopy of Z-DNA. *Nature* 339:484–86.

Avery, O., C. MacLeod, and M. McCarty. 1944. Studies on the chemical nature of the substance inducing transformation of Pneumococcal types. *Journal of Experimental Medicine* 79:137–58.

Baker, T., and S. Bell. 1998. Polymerases and the replisomes: machines within machines. *Cell* 92:295–305.

Baker, T., and S. Wickner. 1992. Genetics and enzymology of DNA replication in *Escherichia coli. Annual Review of Genetics* 26:447–77.

Berger, J., et al. 1996. Structure and mechanism of DNA topoisomerase II. *Nature* 379:225–32.

Cairns, J. 1963. The chromosomes of *E. coli. Cold Spring Harbor Symposium on Quantitative Biology* 28:43–46.

Cann, I., and Y. Ishino. 1999. Archaeal DNA replication: Identifying the pieces to solve a puzzle. *Genetics* 152:1249–67.

Chargaff, E., and J. Davidson, eds. 1955. *The Nucleic Acids.* New York: Academic Press.

Clayton, D. 1991. Replication and transcription of vertebrate mitochondrial DNA. *Annual Review of Cell Biology* 7:453–78.

Dickerson, R., et al. 1982. The anatomy of A-, B-, and Z-DNA. *Science* 216:475–85.

Donovan, S., and J. Diffley. 1996. Replication origins in eukaryotes. *Current Opinion in Genetics and Development* 6:203–7.

Driscoll, R., M. Youngquist, and D. Baldeschwieler. 1990. Atomic-scale imaging of DNA using scanning tunneling microscopy. *Nature* 346:294–96.

Falkenberg, M, I. Lehman, and P. Elias. 2000. Leading- and lagging-strand DNA synthesis *in vitro* by a reconstituted herpes simplex virus type 1 replisome. *Proceedings of the National Academy of Sciences, USA* 97:3896–3900.

Fraenkel-Conrat, H., and B. Singer. 1957. Virus reconstitution. II: Combination of protein and nucleic acid from different strains. *Biochimica et Biophysica Acta* 24:540–48.

Frank-Kamenetskii, M., and S. Mirkin. 1995. Triplex DNA structures. *Annual Review of Biochemistry* 64:65–95.

Griffith, F. 1928. Significance of pneumococcal types. *Journal of Hygiene* 27:113–59.

Hershey, A., and M. Chase. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* 36:39–56.

Holt, I., H. Lorimer, and H. Jacobs. 2000. Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA. *Cell* 100:515–24.

Hübscher, U., H. Nasheuer, and J. Syväoja. 2000. Eukaryotic DNA polymerases, a growing family. *Trends in Biochemical Sciences* 25:143–47.

Jacobs, C., and L. Shapiro. 1999. Bacterial cell division: A moveable feast. *Proceedings of the National Academy of Sciences, USA* 96:5891–93.

Joyce, C., and T. Steitz. 1994. Function and structure relationships in DNA polymerases. *Annual Review of Biochemistry* 63:777–822.

Kamada, K., et al. 1996. Structure of a replication-terminator protein complexed with DNA. *Nature* 383:598–603.

Keck, J., and J. Berger. 1999. Enzymes that push DNA around. *Nature Structural Biology* 6:900–902.

Kelman, Z., and M. O'Donnell. 1995. DNA polymerase III holoenzyme: Structure and function of a chromosomal replicating machine. *Annual Review of Biochemistry* 64:171–200.

Kornberg, A. 1989. *For the Love of Enzymes: The Odyssey of a Biochemist.* Cambridge, Mass.: Harvard University Press.

Kramer, M., et al. 1998. Lagging-strand replication of rolling-circle plasmids: Specific recognition of the *ssoA*-type origins in different gram-positive bacteria. *Proceedings of the National Academy of Sciences, USA* 95:10505–10.

Lederberg, J. 1991. The gene (H. J. Muller, 1947). *Genetics* 129:313–16.

Lindahl, T., and D. Barnes. 1992. Mammalian DNA ligases. *Annual Review of Biochemistry* 61:251–81.

Lohman, T., and M. Ferrari. 1994. *Escherichia coli* single-stranded DNA-binding protein: Multiple DNA-binding modes and cooperativities. *Annual Review of Biochemistry* 63:527–70.

Manna, A., et al. 1996. The dimer-dimer interaction surface of the replication terminator protein of *Bacillus subtilis* and termination of DNA replication. *Proceedings of the National Academy of Sciences, USA* 93:3253–58.

———. 1996. Helicase-contrahelicase interaction and the mechanism of termination of DNA replication. *Cell* 87:881–91.

Marahens, Y., and B. Stillman. 1992. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* 255:817–23.

Marians, K. 1992. Prokaryotic DNA replication. *Annual Review of Biochemistry* 61:673–719.

Marmur, J., and P. Doty. 1962. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *Journal of Molecular Biology* 5:109–18.

Marx, J. 1995. How DNA replication originates. *Science* 270:1585–87.

Meselson, M., and F. Stahl. 1958. The replication of DNA in *Escherichia coli. Proceedings of the National Academy of Sciences, USA* 44:671–82.

Moon, K., et al. 1999. Identification and reconstitution of the origin of recognition complex from *Schizosaccharomyces pombe. Proceedings of the National Academy of Sciences, USA* 96:12367–72.

Morais-Cabral, J., et al. 1997. Crystal structure of the breakage-reunion domain of DNA gyrase. *Nature* 388:903–6.

Nelson, J., C. Lawrence, and D. Hinkle. 1996. Thymine-thymine dimer bypass by yeast DNA polymerase ζ. *Science* 272:1646–49.

Niki, H., and S. Hiraga. 1998. Polar localization of the replication origin and terminus in *Escherichia coli* nucleoids during chromosome partitioning. *Genes & Development* 12:1036–45.

Novick, R. 1998. Contrasting lifestyles of rolling-circle phages and plasmids. *Trends in Biochemical Sciences* 23:434–38.

Prusiner, S., ed. 1999. *Prion Biology and Diseases.* Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press.

Rangarajan, S., et al. 1997. *Escherichia coli* DNA polymerase II catalyzes chromosomal and episomal DNA synthesis *in vivo. Proceedings of the National Academy of Sciences, USA* 94:946–51.

Rich, A., A. Nordheim, and A. Wang. 1984. The chemistry of left-handed Z-DNA. *Annual Review of Biochemistry* 53:791–846.

Roca, J. 1995. The mechanisms of DNA topoisomerases. *Trends in Biochemical Sciences* 20:156–60.

Rothfield, L., S. Justice, and J. Garcia-Lara. 1999. Bacterial cell division. *Annual Review of Genetics* 33:423–48.

Shamoo, Y., and T. Steitz. 1999. Building a replisomes from interacting pieces: Sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* 99:155–66.

Shapiro, L., and R. Losick. 2000. Dynamic spatial regulation in the bacterial cell. *Cell* 100:89–98.

Sharpe, M., and J. Errington. 1999. Upheaval in the bacterial nucleoid. *Trends in Genetics* 15:70–74.

Stillman, B. 1994. Smart machines at the DNA replication fork. *Cell* 78:725–28.

Strobel, S., and P. Dervan. 1990. Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-helix formation. *Science* 249:73–75.

Subramanya, H., et al. 1996. Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell* 85:607–15.

Tessman, I., and M. Kennedy. 1994. DNA polymerase II of *Escherichia coli* in the bypass of abasic sites *in vivo. Genetics* 136:439–48.

Toyn, J., et al. 1995. The activation of DNA replication in yeast. *Trends in Biochemical Sciences* 20:70–73.

Turchi, J., et al. 1994. Enzymatic completion of mammalian lagging-strand DNA replication. *Proceedings of the National Academy of Sciences, USA* 91:9803–7.

Vlieghe, D., et al. 1996. Parallel and antiparallel $(G \bullet GC)_2$ triple helix fragments in a crystal structure. *Science* 273: 1702–5.

Waga, S., and B. Stillman. 1994. Anatomy of a DNA replication fork revealed by a reconstitution of SV40 DNA replication in vitro. *Nature* 369:207–12.

———. 1998. The DNA replication fork in eukaryotic cells. *Annual Review of Biochemistry* 67:721–51.

Wang, A., et al. 1981. Left-handed double helical DNA: Variations in the backbone conformation. *Science* 211:171–76.

Wang, J. 1996. DNA topoisomerases. *Annual Review of Biochemistry* 65:635–92.

Watson, J. 1968. *The Double Helix.* New York: Signet.

Watson, J., and F. Crick. 1953. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* 171:737–38.

West, S. 1996. DNA helicase: New breeds of translocating motors and molecular pumps. *Cell* 86:177–80.

Westheimer, F. 1987. Why nature chose phosphates. *Science* 235:1173–78.

Wheeler, R., and L. Shapiro. 1997. Bacterial chromosome segregation: Is there a mitotic apparatus? *Cell* 88:577–79.

Wickner, R. 1996. Prions and RNA viruses of *Saccharomyces cerevisiae. Annual Review of Genetics* 30:109–139.

Yuzhakov, A., Z. Kelman, and M. O'Donnell. 1999. Trading places on DNA—A three-point switch underlies primer handoff from primase to the replicative DNA polymerase. *Cell* 96:153–63.

Zubay, G. 1997. *Biochemistry,* 4th ed. Dubuque, Iowa: Wm. C. Brown Publishers.

### Chapter 10 Gene Expression: Transcription

Arkhipova, I. 1995. Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics* 139: 1359–69.

Belfort, M. 1990. Phage T4 introns: Self-splicing and mobility. *Annual Review of Genetics* 24:363–85.

Bell, S., et al. 1999. Orientation of the transcription preinitiation complex in Archaea. *Proceedings of the National Academy of Sciences, USA* 96:13662–67.

Bentley, D. 1999. Coupling RNA polymerase II transcription with pre-mRNA processing. *Current Opinion in Cell Biology* 11:347–51.

Brewer, B. 1988. When polymerases collide: Replication and the transcriptional organization of the *E. coli* chromosome. *Cell* 53:679–86.

Brody, E., and J. Abelson. 1985. The "spliceosome": Yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* 228:963–67.

Burley, S., and R. Roeder. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annual Review of Biochemistry* 65:769–99.

Busby, S., and R. Ebright. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79:743–46.

Cattaneo, R. 1991. Different types of messenger RNA editing. *Annual Review of Genetics* 28:71–88.

Cech, T. 1986. RNA as an enzyme. *Scientific American,* November, 64–75.

———. 1990. Self-splicing of group I introns. *Annual Review of Biochemistry* 59:543–68.

Cech, T., and O. Uhlenbeck. 1994. Hammerhead nailed down. *Nature* 372:39–40.

Cheng, S.-W., et al. 1991. Functional importance of sequence in the stem-loop of a transcription terminator. *Science* 254:1205–7.

Conaway, J., and R. Conaway. 1999. Transcription elongation and human disease. *Annual Review of Biochemistry* 68:301–19.

Cramer, P., et al. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288:640–49.

Crick, F. 1970. Central dogma of molecular biology. *Nature* 227:561–63.

Das, A. 1993. Control of transcription termination by RNA-binding proteins. *Annual Review of Biochemistry* 62:893–930.

Di Nicola Negri, E., et al. 1997. The eucaryal tRNA splicing endonuclease recognizes a tripartite set of RNA elements. *Cell* 89:859–66.

Dinter-Gottlieb, G. 1986. Viroids and virusoids are related to group I introns. *Proceedings of the National Academy of Sciences, USA* 83:6250–54.

Dombroski, A., et al. 1996. The sigma subunit of *Escherichia coli* RNA polymerase senses promoter spacing. *Proceedings of the National Academy of Sciences, USA* 93:8858–62.

Dorit, R., L. Schoenbach, and W. Gilbert. 1990. How big is the universe of exons? *Science* 250:1377–82.

Dove, S., J. Joung, and A. Hochschild. 1997. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* 386:627–30.

Eckstein, F., and D. Lilley, eds. 1997. *Catalytic RNA.* New York: Springer.

Estrem, S., et al. 1999. Bacterial promoter architecture: Subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase subunit. *Genes & Development* 13:2134–47.

Ferat, J.-L., and F. Michel. 1993. Group II self-splicing introns in bacteria. *Nature* 364:358–61.

Frank, D., and N. Pace. 1998. Ribonuclease P: Unity and diversity in a tRNA-processing ribozyme. *Annual Review of Biochemistry* 67:153–80.

Fu, J., et al. 1999. Yeast RNA polymerase II at 5 Å resolution. *Cell* 98:799–810.

Garrett, R. 1999. Mechanics of the ribosome. *Nature* 400:811–12. (Introducing two crystallographic studies.)

Geiselmann, J., et al. 1993. A physical model for the translocation and helicase activities of *Escherichia coli* transcription termination protein rho. *Proceedings of the National Academy of Sciences, USA* 90:7754–58.

Gesteland, R., T. Cech, and J. Atkins, eds. 1999. *The RNA World,* 2d ed. Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press.

Goodrich, J., G. Cutler, and R. Tjian. 1996. Contacts in context: Promoter specificity and macromolecular interactions in transcription. *Cell* 84:825–30.

Guthold, M., et al. 1994. Following the assembly of RNA polymerase-DNA complexes in aqueous solutions with the scanning force microscope. *Proceedings of the National Academy of Sciences, USA* 91:12927–31.

Hahn, S. 1998. The role of TAFs in RNA polymerase II transcription. *Cell* 95:579–82.

Hall, B., and S. Spiegelman. 1961. Sequence complementarity of T2-DNA and T2-specific RNA. *Proceedings of the National Academy of Sciences, USA* 47:137–46.

Helmann, J., and M. Chamberlain. 1988. Structure and function of bacterial sigma factors. *Annual Review of Biochemistry* 57:839–72.

Holley, R., et al. 1965. Structure of a ribonucleic acid. *Science* 147: 1462–65.

Kable, M., et al. 1996. RNA editing: A mechanism for gRNA-specified uridylate insertion into precursor mRNA. *Science* 273:1189–95.

Kim, Y., et al. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365:512–20.

Koleske, A., and R. Young. 1995. The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends in Biochemical Sciences* 20:113–17.

Krämer, A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annual Review of Biochemistry* 65:367–409.

Lake, J. 1976. Ribosome structure determined by electron microscopy of *Escherichia coli* small subunits, large subunits and monomeric ribosomes. *Journal of Molecular Biology* 105:131–59.

Liu, B., and B. Alberts. 1995. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* 267:1131–37.

Liu, B., et al. 1993. The DNA replication fork can pass RNA polymerase without displacing the nascent transcript. *Nature* 366:33–39.

Lopez, A. 1998. Alternative splicing of pre-mRNA:Developmental consequences and mechanisms of regulation. *Annual Review of Genetics* 32:279–305.

Lykke-Anderson, J., et al. 1997. Archaeal introns: Splicing, intercellular mobility and evolution. *Trends in Biochemical Sciences* 22:326–31.

Maréchal-Drouard, L., J. Weil, and A. Dietrich. 1993. Transfer RNAs and transfer RNA genes in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 44:13–32.

Marmorstein, R., et al. 1992. DNA recognition by GAL4: Structure of a protein-DNA complex. *Nature* 356:408–14.

McCarthy, B., and J. Holland. 1965. Denatured DNA as a direct template for in vitro protein synthesis. *Proceedings of the National Academy of Sciences, USA* 54:880–86.

McCracken, S., et al. 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385:357–61.

McKeown, M. 1992. Alternative mRNA splicing. *Annual Review of Cell Biology* 8:133–55.

Michel, F., and J.-L. Ferat. 1995. Structure and activities of group II introns. *Annual Review of Biochemistry* 64:435–61.

Miller, O., Jr., et al. 1970. Electron microscopic visualization of transcription. *Cold Spring Harbor Symposium on Quantitative Biology* 35:505–12.

Mistell, T., J. Caceres, and D. Spector. 1997. The dynamics of pre-mRNA splicing factor in living cells. *Nature* 387:523–27.

Murray, H., and K. Jarrell. 1999. Flipping the switch to an active spliceosome. *Cell* 96:599–602.

Naryshkin, N., et al. 2000. Structural organization of the RNA polymerase-promoter open complex. *Cell* 101:601–11.

Nikolov, D., and S. Burley. 1997. RNA polymerase II transcription initiation: a structural view. *Proceedings of the National Academy of Sciences, USA* 94:15–22.

Olson, M., M. Dundr, and A. Szebeni. 2000. The nucleolus: An old factory with unexpected capabilities. *Trends in Cell Biology* 10:189–96.

Pabo, C., and R. Sauer. 1992. Transcription factors: Structural families and principles of DNA recognition. *Annual Review of Biochemistry* 61:1053–95.

Peterson, M., et al. 1991. Structure and functional properties of human general transcription factor IIE. *Nature* 354:369–73.

Pley, H., K. Flaherty, and D. McKay. 1994. Three-dimensional structure of a hammerhead ribozyme. *Nature* 372:68–74.

Proudfoot, N. 1996. Ending the message is not so simple. *Cell* 87:779–81.

Ptashne, M., and A. Gann. 1997. Transcriptional activation by recruitment. *Nature* 386:569–77.

Pyle, A. 1993. Ribozymes: A distinct class of metalloenzymes. *Science* 261:709–14.

Rich, A. 1993. The structure and biology of transfer RNA. *Structure* 1:vi–vii.

Riesner, D., and H. Gross. 1985. Viroids. *Annual Review of Biochemistry* 54:531–64.

Roeder, R. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends in Biochemical Sciences* 21:327–35.

Ross, W., et al. 1993. A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase. *Science* 262:1407–13.

Schafer, D., et al. 1991. Transcription by single molecules of RNA polymerase observed by light microscopy. *Nature* 352:444–48.

Scheer, U., and R. Hock. 1999. Structure and function of the nucleolus. *Current Opinion in Cell Biology* 11:385–90.

Scott, W., and A. Klug. 1996. Ribozymes: Structure and mechanism in RNA catalysis. *Trends in Biochemical Sciences* 21:220–24.

Sharp, P. 1994. Split genes and RNA splicing. *Cell* 77:805–15.

Shih, M.-C., P. Heinrich, and H. Goodman. 1988. Intron existence predated the divergence of eukaryotes and prokaryotes. *Science* 242:1164–66.

Shimamoto, A., et al. 1996. A unique human gene that spans over 230 kb in the human chromosome 8p11-12 and codes multiple family proteins sharing RNA-binding motifs. *Proceedings of the National Academy of Sciences, USA* 93: 10913–17.

Sleeman, J., and A. Lamond. 1999. Nuclear organization of pre-mRNA splicing factors. *Current Opinion in Cell Biology* 11:372–77.

Smith, C., and J. Steitz. 1997. Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell* 89:669–72.

Smith, C., J. Patton, and B. Nadal-Ginard. 1989. Alternative splicing in the control of gene expression. *Annual Review of Genetics* 23:527–77.

Steinmetz, E., and T. Platt. 1994. Evidence supporting a tethered tracking model for helicase activity of *Escherichia coli* rho factor. *Proceedings of the National Academy of Sciences. USA* 91:1401–5.

Steitz, J. 1988. "Snurps." *Scientific American,* June, 56–63.

Stillman, B., ed. 1998. *Mechanisms of Transcription.* Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press.

Stoltzfus, A., et al. 1994. Testing the exon theory of genes: The evidence from protein structure. *Science* 265:202–7.

Tarn, W.-Y., and J. Steitz. 1997. Pre-mRNA splicing: The discovery of a new spliceosome doubles the challenge. *Trends in Biochemical Sciences* 22:132–37.

Tennyson, C., H. Klamut, and R. Worton. 1995. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nature Genetics* 9:184–90.

Tjian, R. 1995. Molecular machines that control genes. *Scientific American,* February, 54–61.

Tjian, R., and T. Maniatis. 1994. Transcriptional activation: A complex puzzle with few easy pieces. *Cell* 77:5–8.

Uptain, S., C. Kane, and M. Chamberlin. 1997. Basic mechanisms of transcript elongation and its regulation. *Annual Review of Biochemistry* 66:117–72.

Venema, J., and D. Tollervey. 1999. Ribosome synthesis in *Saccharomyces cerevisiae. Annual Review of Genetics* 33: 261–311.

Von Hippel, P. 1994. Passing lanes for polymerases? *Current Biology* 4:333–36.

Waldrop, M. 1992. Finding RNA makes proteins gives "RNA world" a big boost. *Science* 256:1396–97.

Wang, J.-F., and T. Cech. 1992. Tertiary structure around the guanosine-binding site of the *Tetrahymena* ribozyme. *Science* 256:526–29.

Wedekind, J., and D. McKay. 1998. Crystallographic structures of the hammerhead ribozyme: Relationship to ribozyme folding and catalysis. *Annual Review of Biophysics and Biomolecular Structure* 27:475–502.

White, R. 1998. *RNA Polymerase III Transcription.* New York: Springer.

Wieczorek, E., et al. 1998. Function of $\text{TAF}_{\text{II}}$-containing complex without TBP in transcription by RNA polymerase II. *Nature* 393:187–91.

Will, C., et al. 1999. Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* 284:2003–5.

Wolin, S., and A. Matera. 1999. The trials and travels of tRNA. *Genes & Development* 13:1–10.

Woodson, S., and N. Leontis. 1998. Structure and dynamics of ribosomal RNA. *Current Opinion in Structural Biology* 8:294–300.

Zaug, A., and T. Cech. 1986. The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science* 231:470–75.

Zawel, L., and D. Reinberg. 1995. Common themes in assembly and function of eukaryotic transcription complexes. *Annual Review of Biochemistry* 64:533–61.

Zeigler, D., and D. Dean. 1990. Orientation of genes in the *Bacillus subtilis* chromosome. *Genetics* 125:703–8.

Zorio, D., et al. 1994. Operons as a common form of chromosomal organization in *C. elegans. Nature* 372:270–72.

### Chapter 11 Gene Expression: Translation

Arnez, J., and D. Moras. 1997. Structural and functional considerations of the aminoacylation reaction. *Trends in Biochemical Sciences* 22:211–16.

Batey, R., et al. 2000. Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* 287:1232–39.

Berchtold, H., et al. 1993. Crystal structure of active elongation factor Tu reveals major domain rearrangements. *Nature* 365:126–32.

Bernabeu, C., and J. Lake. 1982. Nascent polypeptide chains emerge from the exit domain of the large ribosomal subunit: Immune mapping of the nascent chains. *Proceedings of the National Academy of Sciences, USA* 79:3111–15.

Boyd, L., and C. Thummel. 1993. Selection of CUG and AUG initiator codons for *Drosophila E74A* translation depends on downstream sequences. *Proceedings of the National Academy of Sciences, USA* 90:9164–67.

Bukau, B., et al. 2000, Getting newly synthesized proteins into shape. *Cell* 101:119–22.

Capel, M., et al. 1987. A complete mapping of the proteins in the small ribosomal subunit of *Escherichia coli. Science* 238:1403–6.

Caras, I., and G. Weddell. 1989. Signal peptide for protein secretion directing glycophospholipid membrane anchor attachment. *Science* 243:1196–98.

Cate, J., et al. 1999. X-ray crystal structures of 70S ribosome functional complexes. *Science* 285:2095–104.

Chapeville, F., et al. 1962. On the role of soluble ribonucleic acid in coding for amino acids. *Proceedings of the National Academy of Sciences, USA* 48:1086–92.

Chen, C.-Y., and P. Sarnow. 1995. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* 268:415–17.

Chin, G., and J. Marx, eds. 1994. Resistance to antibiotics: Frontiers in biotechnology. *Science* 264:359–93.

Chothia, C., and A. Finkelstein. 1990. The classification and origins of protein folding patterns. *Annual Review of Biochemistry* 59:1007–39.

Cigan, A., L. Feng, and T. Donahue. 1988. $\text{tRNA}_i^{\text{Met}}$ functions in directing the scanning ribosome to the start site of translation. *Science* 242:93–97.

Crick, F. 1966. Codon-anticodon pairing: The wobble hypothesis. *Journal of Molecular Biology* 19:548–55.

Curnow, A., et al. 1997. Glu-$\text{tRNA}^{\text{Gln}}$ amidotransferase: A novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proceedings of the National Academy of Sciences, USA* 94:11819–26.

Ellis, R. 1996. *The Chaperonins.* San Diego: Academic Press.

Feltham, J., and L. Gierasch. 2000. GroEL-substrate interactions: Molding the fold, or folding the mold? *Cell* 100:193–96.

Frank, J. 1998. How the ribosome works. *American Scientist* 86:428–39.

Freymann, D., et al. 1997. Structure of the conserved GTPase domain of the signal recognition particle. *Nature* 385:361–64.

Gabashvilli, I., et al. 2000. Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell* 100:537–49.

Gerber, A., and W. Keller. 1999. An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* 286:1146–49.

Gingras, A., B. Raught, and N. Sonenberg. 1999. eIF4 initiation factors: Effectors of mRNA recruitment to ribosomes and regulators of translation. *Annual Review of Biochemistry* 68:913–63.

Green, R., and H. Noller. 1997. Ribosomes and translation. *Annual Review of Biochemistry* 66:679–716.

Grunberg-Manago, M. 1963. Polynucleotide phosphorylase. *Progress in Nucleic Acid Research* 1:93–133.

Hendrick, J., and F. Hartl. 1993. Molecular chaperone functions of heat-shock proteins. *Annual Review of Biochemistry* 62:349–84.

Houry, W., et al. 1999. Identification of *in vivo* substrates of the chaperonin GroEL. *Nature* 402:147–54.

Ibba, M., and D. Söll. 1999. Quality control mechanisms during translation. *Science* 286:1893–97.

Ito, K., M. Uno, and Y. Nakamura. 2000. A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature* 403:680–84.

Knight, R., S. Freeland, and L. Landweber. 1999. Selection, history and chemistry: The three faces of the genetic code. *Trends in Biochemical Sciences* 24:241–47.

Kurland, C. 1992. Translational accuracy and the fitness of bacteria. *Annual Review of Genetics* 26:29–50.

Lee, V., G. Miller, and M. Yagisawa. 1999. What's new in the antibiotic pipeline. *Current Opinion in Microbiology* 2:475–82.

Levy, S. 1998. The challenge of antibiotic resistance. *Scientific American,* March, 46–53.

Lodmell, J., and A. Dahlberg. 1997. A conformational switch in *Escherichia coli* 16S ribosomal RNA during decoding of messenger RNA. *Science* 277:1262–67.

Low, S., and M. Berry. 1996. Knowing when not to stop: Selenocysteine incorporation in eukaryotes. *Trends in Biochemical Sciences* 21:203–8.

Macejak, D., and P. Sarnow. 1991. Internal initiation of translation mediated by the 5′ leader of a cellular mRNA. *Nature* 353:90–94.

Mande, S., et al. 1996. Structure of the heat shock protein chaperonin-10 of *Mycobacterium leprae. Science* 271:203–7.

McCarthy, J., and R. Brimacombe. 1994. Prokaryotic translation: The interactive pathway leading to initiation. *Trends in Genetics* 10:402–7.

Moazed, D., and H. Noller. 1989. Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell* 57:585–97.

Nakamura, Y., K. Ito, and L. Isaksson. 1996. Emerging understanding of translation termination. *Cell* 87:147–50.

Neupert, W. 1997. Protein import into mitochondria. *Annual Review of Biochemistry* 66:863–917.

Nirenberg, M., and P. Leder. 1964. RNA codewords and protein synthesis: The effect of trinucleotides upon the binding of tRNA to ribosomes. *Science* 145:1399–1407.

Nirenberg, M., and J. Matthei. 1961. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences, USA* 47:1588–1602.

Nissen, P., et al. 1998. The crystal structure of Cys-$\text{tRNA}^{\text{Cys}}$-EF-Tu-GDPNP reveals general and specific features in the ternary complex and in tRNA. *Structure* 7:143–56.

Normanly, J., and J. Abelson. 1989. tRNA identity. *Annual Review of Biochemistry* 58:1029–49.

Ochoa, S. 1980. The pursuit of a hobby. *Annual Review of Biochemistry* 49:1–30.

Preer, J., et al. 1985. Deviation from the universal code shown by the gene for surface protein 51A in Paramecium. *Nature* 314:188–90.

Richards, F. 1991. The protein folding problem. *Scientific American,* January, 54–63.

Rodnina, M., et al. 1997. Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature* 385:37–41.

Ruff, M., et al. 1991. Class II aminoacyl transfer RNA synthetases: Crystal structure of yeast aspartyl-tRNA synthetase complexed with $\text{tRNA}^{\text{Asp}}$. *Science* 252:1682–89.

Ryabova, L., and T. Hohn. 2000. Ribosome shunting in the cauliflower mosaic virus 35S RNA leader is a special case of reinitiation of translation functioning in plant and animal systems. *Genes & Development* 14:817–29.

Sachs, A. 2000. Cell cycle-dependent translation initiation: IRES elements prevail. *Cell* 101:243–45.

Sachs, A., P. Sarnow, and M. Hentze. 1997. Starting at the beginning, middle, and end: Translation initiation in eukaryotes. *Cell* 89:831–38.

Saks, M., J. Sampson, and J. Abelson. 1994. The transfer RNA identity problem: A search for rules. *Science* 263:191–97.

Samaha, R., R. Green, and H. Noller. 1995. A base pair between tRNA and 23S rRNA in the peptidyl transferase centre of the ribosome. *Nature* 377:309–14.

Sanger, F. 1988. Sequences, sequences, and sequences. *Annual Review of Biochemistry* 57:1–28.

Schimmel, P. 1987. Aminoacyl tRNA synthetases: General scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs. *Annual Review of Biochemistry* 56:125–58.

Scott, J., et al. 1983. Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins. *Science* 221:236–40.

Selmer, M., et al. 1999. Crystal structure of *Thermotoga maritima* ribosome recycling factor: a tRNA mimic. *Science* 286:2349–52.

Shtilerman, M., G. Lorimer, and S. Englander. 1999. Chaperonin function: Folding by forced unfolding. *Science* 284:822–25.

Sigler, P., et al. 1998. Structure and function in Gro-EL-mediated protein folding. *Annual Review of Biochemistry* 67:581–608.

Song, H., et al. 2000. The crystal structure of human eukaryotic release factor eRF1—Mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* 100:311–21.

Sprinzl, M. 1994. Elongation factor Tu: A regulatory GTPase with an integrated effector. *Trends in Biochemical Sciences* 19:245–50.

Stadtman, T. 1996. Selenocysteine. *Annual Review of Biochemistry* 65:83–100.

Stansfield, I., and M. Tuite. 1994. Polypeptide chain termination in *Saccharomyces cerevisiae. Current Genetics* 25:385–95.

Stark, H., et al. 1997. Arrangement of tRNAs in pre- and posttranslocational ribosomes revealed by electron cryomicroscopy. *Cell* 88:19–28.

———. 2000. Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation. *Cell* 100:301–9.

Stathopoulos, C., et al. 2000. One polypeptide with two aminoacyl-tRNA synthetase activities. *Science* 287:479–82.

Stern, S., et al. 1989. RNA-protein interactions in 30S ribosomal subunits: Folding and function of 16S rRNA. *Science* 244:783–90.

Vogel, G. 1998. Tracking the history of the genetic code. *Science* 281:329–31.

Wilson, K., and H. Noller. 1998. Molecular movement inside the translational engine. *Cell* 92:337–49.

Wolin, S. 1994. From the elephant to *E. coli*: SRP-dependent protein targeting. *Cell* 77:787–90.

Xu, Z., A. Horwich, and P. Sigler. 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)$_7$ chaperonin complex. *Nature* 388:741–50.

### Chapter 12 DNA: Its Mutation, Repair, and Recombination

Aboussekhra, A., et al. 1995. Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell* 80:859–68.

Ames, B., R. Magaw, and L. Gold. 1987. Ranking possible carcinogenic hazards. *Science* 236:271–80.

Amrein, M., et al. 1988. Scanning tunneling microscopy of RecA-DNA complexes coated with a conducting film. *Science* 240:514–16.

Anderson, D., and S. Kowalczykowski. 1997. The recombination hot spot $\chi$ is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. *Genes & Development* 11:571–81.

Ariyoshi, M., et al. 1994. Atomic structure of the RuvC resolvase: A Holliday junction-specific endonuclease from *E. coli. Cell* 78:1063–72.

Benzer, S. 1961. On the topography of the genetic fine structure. *Proceedings of the National Academy of Sciences, USA* 47:403–15.

Bruner, S., D. Norman, and G. Verdine. 2000. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* 403:859–66.

Bruner, S., et al. 1998. Repair of oxidatively damaged guanine in *Saccharomyces cerevisiae* by an alternative pathway. *Current Biology* 8:393–403.

Buermeyer, A., et al. 1999. Mammalian DNA mismatch repair. *Annual Review of Genetics* 33:533–64.

Cairns, J., J. Overbaugh, and S. Miller. 1988. The origin of mutants. *Nature* 335:142–45.

Camerini-Otero, R., and P. Hsieh. 1995. Homologous recombination proteins in prokaryotes and eukaryotes. *Annual Review of Genetics* 29:509–52.

Cox, M., et al. 2000. The importance of repairing stalled replication forks. *Nature* 404:37–41.

Craig, N. 1988. The mechanism of conservative site-specific recombination. *Annual Review of Genetics* 22:77–105.

Culotta, E., and D. Koshland, Jr. 1994. Molecule of the year: DNA repair works its way to the top. *Science* 266:1926–29.

de Laat, W., N. Jaspers, and J. Hoeijmakers. 1999. Molecular mechanisms of nucleotide excision repair. *Genes & Development* 13:768–85.

Drake, J., et al. 1998. Rates of spontaneous mutation. *Genetics* 148:1667–86.

Eggleston, A., and S. West. 1996. Exchanging partners: recombination in *E. coli. Trends in Genetics* 12:20–26.

Epstein, R., et al. 1963. Physiological studies of conditional lethal mutations of bacteriophage T4D. *Cold Spring Harbor Symposium on Quantitative Biology* 28:375–92.

Featherstone, C., and S. Jackson. 1999. Ku, a DNA repair protein with multiple cellular functions? *Mutation Research* 434:3–15.

Foster, P. 1999. Mechanisms of stationary phase mutation: A decade of adaptive mutation. *Annual Review of Genetics* 33:57–88.

Friedberg, E. 1996. Relationships between DNA repair and transcription. *Annual Review of Biochemistry* 65:15–42.

Friedberg, E., G. Walker, and W. Siede. 1995. *DNA Repair and Mutagenesis.* Herndon, Va.: ASM Press.

Goodman, M., and B. Tippin. 2000. Sloppier copier DNA polymerases involved in genome repair. *Current Opinion in Genetics & Development* 10:162–68.

Green, M. 1990. The foundations of genetic fine structure: A retrospective from memory. *Genetics* 124:793–96.

Hall, B. 1997. On the specificity of adaptive mutations. *Genetics* 145:39–44.

Hanawalt, P., B. Donahue, and K. Sweder. 1994. Repair and transcription: Collision or collusion? *Current Biology* 4:518–21.

Hoeijmakers, J. 1993. Nucleotide excision repair I: From *E. coli* to yeast. *Trends in Genetics* 9:173–77.

———. 1993. Nucleotide excision repair II: From yeast to mammals. *Trends in Genetics* 9:211–17.

Holliday, R. 1974. Molecular aspects of genetic exchange and gene conversion. *Genetics* 78:273–87.

Hopfner, K.-P., et al. 2000. Structural biology of the Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* 101:789–800.

Johnson, R., S. Prakash, and L. Prakash. 1999. Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Polη. *Science* 283:1001–4.

Kanaar, R., J. Hoeijmakers, and D. van Gent. 1998. Molecular mechanisms of DNA double-strand break repair. *Trends in Cell Biology* 8:483–89.

Karran, P. 2000. DNA double-strand break repair in mammalian cells. *Current Opinion in Genetics & Development* 10:144–50.

Keeney, S., C. Giroux, and N. Kleckner. 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* 88:375–84.

Kobayashi, T., et al. 1997. Mutations in the *XPD* gene leading to xeroderma pigmentosum symptoms. *Human Mutation* 9:322–31.

Kowalczykowski, S., and A. Eggleston. 1994. Homologous pairing and DNA strand-exchange proteins. *Annual Review of Biochemistry* 63:991–1043.

Lee, C., and M. Levitt. 1991. Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core. *Nature* 352:448–51.

Ley, R. 1993. Photoreactivation in humans. *Proceedings of the National Academy of Sciences, USA* 90:4337.

Lindahl, T., and R. Wood. 1999. Quality control by DNA repair. *Science* 286:1897–1905.
Luria, S., and M. Delbrück. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28:491–511.
McCullough, A, M. Dodson, and R. Lloyd. 1999. Initiation of base excision repair: Glycosylase mechanisms and structures. *Annual Review of Biochemistry* 68:255–85.
Modrich, P., and R. Lahue. 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annual Review of Biochemistry* 65:101–33.
Mol, C., et al. 1999. DNA repair mechanisms for the recognition and removal of damaged DNA bases. *Annual Review of Biophysics and Biomolecular Structure* 28:101–28.
Morgan, A. 1993. Base mismatches and mutagenesis: How important is tautomerism? *Trends in Biochemical Sciences* 18:160–63.
Muller, H. 1927. Artificial transmutation of the gene. *Science* 66:84–87.
Neel, J. 1994. *Physician to the Gene Pool: Genetic Lessons and Other Stories.* New York: Wiley.
Ortiz-Lombardia, M., et al. 1999. Crystal structure of a DNA Holliday junction. *Nature Structural Biology* 6:913–17.
Parikh, S., C. Mol, and J. Tainer. 1997. Base excision repair enzyme family portrait: Integrating the structure and chemistry of an entire DNA repair pathway. *Structure* 5:1543–50.
Potter, H., and D. Dressler. 1976. On the mechanism of genetic recombination: Electron microscopic observation of recombination intermediates. *Proceedings of the National Academy of Sciences, USA* 73:3000–3004.
Puchta, H., and B. Hohn. 1996. From centimorgans to base pairs: homologous recombination in plants. *Trends in Plant Science* 1:340–48.
Rafferty, J., et al. 1996. Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science* 274:415–21.
Richardson, C., and M. Jasin. 2000. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* 405:697–700.
Ripley, L. 1990. Frameshift mutation: Determinants of specificity. *Annual Review of Genetics* 24:189–213.
Roberts, R., and X. Cheng. 1998. Base flipping. *Annual Review of Biochemistry* 67:181–98.
Roman, H., and M. Ruzinski. 1990. Mechanisms of gene conversion in *Saccharomyces cerevisiae. Genetics* 124:7–25.
Rosche, W., P. Foster, and J. Cairns. 1999. The role of transient hypermutators in adaptive mutation in *Escherichia coli. Proceedings of the National Academy of Sciences, USA* 96:6862–67.
Sancar, A. 1996. DNA excision repair. *Annual Review of Biochemistry* 65:43–81.
Singer, B., and J. Kusmierak. 1982. Chemical mutagenesis. *Annual Review of Biochemistry* 52:655–93.
Smith, M. 1985. In vitro mutagenesis. *Annual Review of Genetics* 19:423–62.
Stadler, L. 1928. Mutations in barley induced by X rays and radium. *Science* 68:186–87.
Stahl, F. 1994. The Holliday junction on its thirtieth anniversary. *Genetics* 138:241–46.
———. 1996. Meiotic recombination in yeast: coronation of the double-strand-break repair model. *Cell* 87:965–68.
Story, R., I. Weber, and T. Steitz. 1992. The structure of *E. coli* RecA protein monomer and polymer. *Nature* 355:318–25 (Erratum, p.567).
Streisinger, G., and J. Owen. 1985. Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4. *Genetics* 109:633–59.
Tanaka, K., and R. Wood. 1994. Xeroderma pigmentosum and nucleotide excision repair of DNA. *Trends in Biochemical Sciences* 19:83–86.
Tang, M., et al. 2000. Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature* 404:1014–18.
Tennant, R., et al. 1987. Prediction of chemical carcinogenicity in rodents from in vitro genetic toxicity assays. *Science* 236:933–41.
Van Dyck, E., et al. 1999. Binding of double-strand breaks in DNA by human Rad52 protein. *Nature* 398:728–31.
Washington, M., et al. 2000. Accuracy of thymine-thymine dimer bypass by *Saccharomyces cerevisiae* DNA polymerase $\eta$. *Proceedings of the National Academy of Sciences, USA* 97:3094–99.
West, S. 1994. The processing of recombination intermediates: Mechanistic insights from studies of bacterial proteins. *Cell* 76:9–15.
———. 1997. Processing of recombination intermediates by the RuvABC proteins. *Annual Review of Genetics* 31:213–44.
Wood, R. 1996. DNA repair in eukaryotes. *Annual Review of Biochemistry* 65:135–67.
Yanofsky, C., et al. 1967. The complete amino-acid sequence of the tryptophan synthetase A protein (a subunit) and its colinear relationship with the genetic map of the A gene. *Proceedings of the National Academy of Sciences, USA* 57:296–98.

### Chapter 13 Genomics, Biotechnology, and Recombinant DNA

Abelson, P., and P. Hines. 1999. The plant revolution. *Science* 285:367–68. (Followed by seven articles on plant genomics.)
Adams, M., et al. 2000. The genome sequence of *Drosophila melanogaster. Science* 287:2185–95. (Followed by seven related articles.)
Adelman, J., et al. 1987. Two mammalian genes transcribed from opposite strands of the same DNA locus. *Science* 235:1514–17.
Andreason, G., and G. Evans. 1988. Introduction and expression of DNA molecules in eukaryotic cells by electroporation. *BioTechniques* 6:650–60.
Armour, J., and A. Jeffreys. 1992. Biology and applications of human minisatellite loci. *Current Opinion in Genetics and Development* 2:850–56.
Arnheim, N., and H. Erlich. 1992. Polymerase chain reaction strategy. *Annual Review of Biochemistry* 61:131–56.
Ayala, F., and B. Black. 1993. Science and the courts. *American Scientist* 81:230–39.
Berg, P., and M. Singer. 1995. The recombinant DNA controversy: Twenty years later. *Proceedings of the National Academy of Sciences, USA* 92:9011–13.
Blattner, F., et al. 1997. The complete genome sequence of *Escherichia coli* K12. *Science* 277:1453–62.
Bolivar, F., et al. 1977. Construction and characterization of new cloning vehicles. II: A multipurpose cloning system. *Gene* 2:95–113.
Brandon, E., R. Idzerda, and G. McKnight. 1995. Targeting the mouse genome: A compendium of knockouts. *Current Biology,* 5:625–34 (part I); 5:758–65 (part II); 5:873–81 (part III).
Cai, W., et al. High-resolution restriction maps of bacterial artificial chromosomes constructed by optical mapping. *Proceedings of the National Academy of Sciences, USA* 95:3390–95.
Capecchi, M. 1994. Targeted gene replacement. *Scientific American,* March, 52–59.
Caruthers, M. 1985. Gene synthesis machines: DNA chemistry and its uses. *Science* 230:281–85.
*C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans:* A platform for investigating biology. *Science* 282:2012–18. (Followed by five articles.)
Chu, T.-H., et al. 1995. Highly efficient eukaryotic gene expression vectors for peptide secretion. *BioTechniques* 18:890–99.
Cohen, J., and M. Hogan. 1994. The new genetic medicines. *Scientific American,* December, 76–82.

Cohen, S., et al. 1973. Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences, USA* 70:3240–44.

Danna, K., and D. Nathans. 1971. Specific cleavage of Simian Virus 40 DNA by restriction endonuclease of *Haemophilus influenzae. Proceedings of the National Academy of Sciences, USA* 68:2913–17.

DePamphilis, M., et al. 1988. Microinjecting DNA into mouse ova to study DNA replication and gene expression and to produce transgenic animals. *BioTechniques* 6:662–80.

Dhand, R., ed. 2000. Functional genomics. *Nature* 405:619–65. (Several articles on genomics and proteomics.)

Dib., C., et al. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152–54.

Dickson, R., et al. 1997. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* 388:355–58.

Dunham, I., et al. 1999. The DNA sequence of human chromosome 22. *Nature* 402:489–95.

Elder, J., R. Spritz, and S. Weissman. 1981. Simian virus 40 as a eukaryotic cloning vehicle. *Annual Review of Genetics* 15:295–340.

Ferber, D. 1999. GM crops in the cross hairs. *Science* 286:1662–66.

Fleischmann, R., et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496–512.

Fodor, S. 1997. DNA sequencing: Massively parallel genomics. *Science* 277:393–95.

Friedmann, T., ed. 1999. *The Development of Human Gene Therapy.* Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press.

Gasser, C., and R. Fraley. 1992. Transgenic crops. *Scientific American,* June, 62–69.

Gerhold, D., T. Rushmore, and C. Caskey. 1999. DNA chips: Promising toys have become powerful tools. *Trends in Biochemical Sciences* 24:168–73.

Gheysen, G., G. Angenon, and M. Van Montagu. 1992. Transgenic plants: *Agrobacterium tumefaciens*-mediated transformation and its use for crop improvement. In *Transgenesis,* edited by J. Murray, pp. 187–232. West Sussex, England: John Wiley.

Gilbert, W. 1981. DNA sequencing and gene structure. *Science* 214:1305–12.

Hagelberg, E., I. Gray, and A. Jeffreys. 1991. Identification of the skeletal remains of a murder victim by DNA analysis. *Nature* 352:427–29.

Hall, J., et al. 1990. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250:1684–89.

Hattori, M., et al. 2000. The DNA sequence of human chromosome 21. *Nature* 405:311–19.

Horsch, R., et al. 1985. A simple and general method for transferring genes into plants. *Science* 227:1229–31.

Jasny, B., and P. Hines. 1999. Genome prospecting. *Science* 286:443. (Followed by eight articles.)

Johnson, I. 1983. Human insulin from recombinant DNA technology. *Science* 219:632–37.

Johnston, S., et al. 1988. Mitochondrial transformation in yeast by bombardment with microprojectiles. *Science* 240:1538–41.

Johnston-Dow, L., et al. 1987. Optimized method for fluorescent and radio labeled DNA sequencing. *BioTechniques* 5:754–65.

Kato, Y., et al. 1998. Eight calves cloned from somatic cells of a single adult. *Science* 282:2095.

Kilbane, J., II, and B. Bielaga. 1991. Instantaneous gene transfer from donor to recipient microorganisms via electroporation. *BioTechniques* 10:354–65.

Kmiec, E. 1999. Gene therapy. *American Scientist* 87:240–47.

Kreiner, T. 1996. Rapid genetic sequence analysis using a DNA probe array system. *American Laboratory,* March: 39–43.

Kubota, C., et al. 2000. Six cloned calves produced from adult fibroblast cells after long-term culture. *Proceedings of the National Academy of Sciences, USA* 97:990–95.

Lander, E., and R. Weinberg. 2000. Genomics: Journey to the center of biology. *Science* 287:1777–82.

Lasic, D. 1992. Liposomes. *American Scientist* 80:20–31.

Lim, K., and C.-B. Chae. 1989. A simple assay for DNA transfection by incubation of the cells in culture dishes with substrates for beta-galactosidase. *BioTechniques* 7:576–79.

Malakoff, D. 2000. The rise of the mouse, biomedicine's model mammal. *Science* 288:248–53. (Followed by two related articles.)

Mannino, R., and S. Gould-Fogerite. 1988. Liposome mediated gene transfer. *BioTechniques* 6:682–90.

Marshall, E. 1999. Gene therapy death prompts review of adenovirus vector. *Science* 286:2244–45.

Martin, C., et al. 1991. Improved chemiluminescent DNA sequencing. *BioTechniques* 11:110–13.

Marx, J. 1982. Building bigger mice through gene transfer. *Science* 218:1298.

———. 1988. DNA fingerprinting takes the witness stand. *Science* 240:1616–18.

———. 1989. The cystic fibrosis gene is found. *Science* 245:923–25.

McElfresh, K., D. Vining-Forde, and I. Balazs. 1993. DNA-based identity testing in forensic science. *BioScience* 43:149–57.

McLaren, A. 2000. Cloning: Pathways to a pluripotent future. *Science* 288:1775–80.

Meinke, D., et al. 1998. *Arabidopsis thaliana:* A model plant for genome analysis. *Science* 282:662–82. (Followed by related articles on genomes.)

Miki, Y., et al. 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266:66–71.

Mullis, K. 1990. The unusual origin of the polymerase chain reaction. *Scientific American,* April, 56–65.

Nakamura, Y., et al. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235:1616–22.

Neufeld, P., and N. Colman. 1990. When science takes the witness stand. *Scientific American,* May, 46–53.

Normark, S., et al. 1983. Overlapping genes. *Annual Review of Genetics* 17:499–525.

Nowak, R. 1994. Breast cancer gene offers surprises. *Science* 265:1796–99.

Oliver, S., et al. 1992. The complete DNA sequence of yeast chromosome III. *Nature* 357:38–46.

Ow, D., et al. 1986. Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science* 234:856–59.

Palmiter, R., and R. Brinster. 1986. Germ-line transformation of mice. *Annual Review of Genetics* 20:465–99.

Paoletti, M., and D. Pimentel. 1996. Genetic engineering in agriculture and the environment. *BioScience* 46:665–73.

Pestka, S. 1983. The purification and manufacture of human interferons. *Scientific American,* August, 37–43.

Rennie, J. 1994. Grading the gene tests. *Scientific American,* June, 88–97.

Richa, J., and C. Lo. 1989. Introduction of human DNA into mouse eggs by injection of dissected chromosome fragments. *Science* 245:175–77.

Roberts, L. 1993. Zeroing in on a breast cancer susceptibility gene. *Science* 259:622–25.

Sanger, F. 1988. Sequences, sequences, and sequences. *Annual Review of Biochemistry* 57:1–28.

Sanger, F., et al. 1978. The nucleotide sequence of bacteriophage φX174. *Journal of Molecular Biology* 125:225–46.

Schell, J. 1987. Transgenic plants as tools to study the molecular organization of plant genes. *Science* 237:1176–83.

Schuler, G., et al. 1996. A gene map of the human genome. *Science* 274:540–46.

*Scientific American.* 1997. Special report: Making gene therapy work. June, 95–123 (5 articles).

Shapiro, H. 1997. Ethical and policy issues of human cloning. *Science* 277:195–96.

Shigekawa, K., and W. Dower. 1988. Electroporation of eukaryotes and prokaryotes: A general approach to the introduction of macromolecules into cells. *BioTechniques* 6:742–51.

Snow, A., and P. Palma. 1997. Commercialization of transgenic plants: Potential ecological risks. *BioScience* 47:86–96.

Snyder, R., et al. 1999. Correction of hemophilia B in canine and murine models using recombinant adeno-associated viral vectors. *Nature Medicine* 5:64–70.

Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98:503–17.

Stix, G. 1995. A recombinant feast. *Scientific American,* March, 38–40.

Takada, T., et al. 1997. Selective production of transgenic mice using green fluorescent protein as a marker. *Nature Biotechnology* 15:458–61.

Tanksley, S., et al. 1992. High-density molecular linkage maps of the tomato and potato genomes. *Genetics* 132:1141–60.

Tsien, R. 1998. The green fluorescent protein. *Annual Review of Biochemistry* 67:509–44.

Velander, W., H. Lubon, and W. Drohan. 1997. Transgenic livestock as drug factories. *Scientific American,* January, 70–74.

Venter, J., H. Smith, and L. Hood. 1996. A new strategy for genome sequencing. *Nature* 381:364–66.

Venter, J., et al. 1998. Shotgun sequencing of the human genome. *Science* 280:1540–42.

Wakayama, T., et al. 1998. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* 394:369–74.

Wang, D., et al. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077–82.

White, R., and J. Lalouel. 1988. Chromosome mapping with DNA markers. *Scientific American,* February, 40–48.

Williams, R., et al. 1991. Introduction of foreign genes into tissues of living mice by DNA-coated microprojectiles. *Proceedings of the National Academy of Sciences, USA* 88:2726–30.

Wilmut, I. 1998. Cloning for medicine. *Scientific American,* December, 58–63.

Wilmut, I., et al. 1997. Viable offspring derived from fetal and adult mammalian cells. *Nature* 385:810–13.

Wilson, M., and S. Lindow. 1993. Release of recombinant microorganisms. *Annual Review of Microbiology* 47:913–44.

Wilson, M., et al. 1995. Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. *BioTechniques* 18:662–69.

Wong, W., et al. 1988. Wood hydrolysis by *Cellulomonas fimi* endoglucanase and exoglucanase coexpressed as secreted enzymes in *Saccharomyces cerevisiae. Bio/Technology* 6:713–19.

Wood, K., et al. 1989. Complementary DNA coding click beetle luciferases can elicit bioluminescence of different colors. *Science* 244:700–702.

Ye, X., et al. 2000. Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287:303–5.

Yuan, R. 1981. Structure and mechanism of multifunctional restriction endonucleases. *Annual Review of Biochemistry* 50:285–315.

### Chapter 14 Gene Expression: Control in Prokaryotes and Phages

Albright, R., and B. Matthews. 1998. How Cro and λ-repressor distinguish between operators: The structural basis underlying a genetic switch. *Proceedings of the National Academy of Sciences, USA* 95:3431–36.

Antson, A., et al. 1999. Structure of the *trp* RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* 401:235–42.

Bachmair, A., D. Finley, and A. Varshavsky. 1986. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* 234:179–86.

Beckwith, J., and T. Silvahy. 1992. *The Power of Bacterial Genetics: A Literature-Based Course.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Bell, C., et al. 2000. Crystal structure of the λ repressor C-terminal domain provides a model for cooperative operator binding. *Cell* 101:801–11.

Casjens, S. 1998. The diverse and dynamic structure of bacterial genomes. *Annual Review of Genetics* 32:339–77.

Charlebois, R., ed. 1999. *Organization of the Prokaryotic Genome.* Washington, D.C.: ASM Press.

Chen, Y., Y. Ebright, and R. Ebright. 1994. Identification of the target of a transcription activator protein by protein-protein photocrosslinking. *Science* 265:90–92.

Dickson, R., et al. 1975. Genetic regulation: The *lac* control region. *Science* 187:27–35.

Eguchi, Y., T. Itoh, and J.-I. Tomazawa. 1991. Antisense RNA. *Annual Review of Biochemistry* 60:631–52.

Erie, D.A., et al. 1994. DNA bending by Cro protein in specific and nonspecific complexes: Implications for protein site recognition and specificity. *Science* 266:1562–66.

Gallant, J., and D. Lindsley. 1998. Ribosomes can slide over and beyond "hungry" codons, resuming protein chain elongation many nucleotides downstream. *Proceedings of the National Academy of Sciences, USA* 95:13771–76.

Geiduschek, E. 1991. Regulation of expression of the late genes of bacteriophage T4. *Annual Review of Genetics* 25:437–60.

Ghosh, S., and M. Deutscher. 1999. Oligoribonuclease is an essential component of the mRNA decay pathway. *Proceedings of the National Academy of Sciences, USA* 96:4372–77.

Gilbert, W., and B. Müller-Hill. 1966. Isolation of the *Lac* repressor. *Proceedings of the National Academy of Sciences, USA* 56:1891–98.

Gold, L. 1988. Posttranscriptional regulatory mechanisms in *Escherichia coli. Annual Review of Biochemistry* 57:199–233.

Gottesman, S. 1999. Regulation by proteolysis: Developmental switches. *Current Opinion in Microbiology* 2:142–47.

Greenblatt, J., J. Nodwell, and W. Mason. 1993. Transcriptional antitermination. *Nature* 364:401–5.

Grindley, N. 1997. Site-specific recombination: Synapsis and strand exchange revealed. *Current Biology* 7:R608–12.

Haren, L., B. Ton-Hoang, and M. Chandler. Integrating DNA: Transposases and retroviral integrases. *Annual Review of Microbiology* 53:245–81.

Helmann, J. 1999. Anti-sigma factors. *Current Opinion in Microbiology* 2:135–41.

Holzman, D. 1991. A "jumping gene" caught in the act. *Science* 254:1728–29.

Hughes, K., and K. Mathee. 1998. The anti-sigma factors. *Annual Review of Microbiology* 52:231–86.

Jacob, F. 1988. *The Statue Within: An Autobiography.* Translated from the French by F. Philip, Alfred P. Sloan Foundation Series. New York: Basic Books.

Jacob, F., and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3:318–56.

Johnson, W., C. Moran, Jr., and R. Losick. 1983. Two RNA polymerase sigma factors from *Bacillus subtilis* discriminate between overlapping promoters for a developmentally regulated gene. *Nature* 302:800–804.

Judson, H. 1979. *The Eighth Day of Creation: Makers of the Revolution in Biology.* New York: Simon and Schuster.

Kantrowitz, E., and W. Lipscomb. 1988. *Escherichia coli* aspartate transcarbamylase: The relation between structure and function. *Science* 241:669–74.

Kihara, A., Y. Akiyama, and K. Ito. 1997. Host regulation of lysogenic decision in bacteriophage λ: transmembrane modulation of FtsH (HflB), the cII degrading protease, by HflKC (HflA). *Proceedings of the National Academy of Sciences, USA* 94:5544–49.

Kleckner, N. 1990. Regulating Tn*10* and IS*10* transposition. *Genetics* 124:449–54.

Kolter, R., and C. Yanofsky. 1982. Attenuation in amino acid biosynthetic operons. *Annual Review of Genetics* 16:113–34.

Konigsberg, W., and G. Godson. 1983. Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli. Proceedings of the National Academy of Sciences, USA* 80:687–91.

Kröger, M., and G. Hobom. 1982. Structural analysis of insertion sequence IS*5*. *Nature* 297:159–62.

Lewis, M., et al. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 271:1247–54.

Li, M., H. Moyle, and M. Susskind. 1994. Target of the transcriptional activator function of phage λ cI protein. *Science* 263:75–77.

McGlynn, P., and R. Lloyd. 2000. Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression. *Cell* 101:35–45.

Müller-Hill, B. 1996. *The* lac *Operon.* Berlin: Walter de Gruyter.

Murialdo, H. 1991. Bacteriophage lambda DNA maturation and packaging. *Annual Review of Biochemistry* 60:125–53.

Oxender, D., G. Zurawski, and C. Yanofsky. 1979. Attenuation in the *Escherichia coli* tryptophan operon: Role of RNA secondary structure involving the tryptophan codon region. *Proceedings of the National Academy of Sciences, USA* 76:5524–28.

Ptashne, M. 1992. *A Genetic Switch: Gene Control and Phage* λ, 2d ed. Cambridge, Mass.: Cell Press and Blackwell Scientific Publications.

Rechsteiner, M., and S. Rogers. 1996. PEST sequences and regulation by proteolysis. *Trends in Biochemical Sciences* 21:267–71.

Rogers, S., R. Wells, and M. Rechsteiner. 1986. Amino acid sequences common to rapidly degraded proteins: The PEST hypothesis. *Science* 234:364–68.

Schultz, S., G. Shields, and T. Steitz. 1991. Crystal structure of a CAP-DNA complex: The DNA is bent by 90°. *Science* 253:1001–7.

Sherratt, D. 1995. *Mobile Genetic Elements.* Oxford: Oxford University Press.

Varshavsky, A. 1996. The N-end rule: functions, mysteries, uses. *Proceedings of the National Academy of Sciences, USA* 93:12142–49.

Weintraub, H. 1990. Antisense RNA and DNA. *Scientific American,* January, 40–46.

Yanofsky, C. 1984. Comparison of regulatory and structural regions of genes of tryptophan metabolism. *Molecular Biology and Evolution* 1:143–61.

Yura, T., and K. Nakahigashi. 1999. Regulation of the heat-shock response. *Current Opinion in Microbiology* 2:153–58.

### Chapter 15 The Eukaryotic Chromosome

Bank, A., J. Mears, and F. Ramirez. 1980. Disorders of human hemoglobin. *Science* 207:486–93.

Berendes, H., F. van Breugel, and T. Holt. 1965. Experimental puffs in salivary gland chromosomes of *Drosophila hydei.* Chromosoma (Berlin) 16:35–46.

Berger, S. 1999. Gene activation by histone and factor acetyltransferases. *Current Opinion in Cell Biology* 11:336–41.

Blackburn, E. 1992. Telomerases. *Annual Review of Biochemistry* 61:113–29.

Bonne-Andrea, C., M. Wong, and B. Alberts. 1990. In vitro replication through nucleosomes without histone displacement. *Nature* 343:719–26.

Britten, R., and D. Kohne. 1968. Repeated sequences in DNA. *Science* 161:529–40.

Broda, P., S. Oliver, and P. Sims, eds. 1993. The eukaryotic genome. *Society for General Microbiology,* Symposium 50:1–407.

Callan, H. 1986. *Lampbrush Chromosomes, vol. 36: Molecular Biology, Biochemistry, and Biophysics.* New York: Springer-Verlag.

Clarke, L., and J. Carbon. 1985. The structure and function of yeast centromeres. *Annual Review of Genetics* 19:29–56.

Collins, K. 1999. Ciliate telomerase biochemistry. *Annual Review of Biochemistry* 68:187–218.

Comings, D. 1978. Mechanisms of chromosome banding and implications for chromosome structure. *Annual Review of Genetics* 12:25–46.

Copenhaver, G., et al. 1999. Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* 286:2468–74.

Costanzi, C., and J. Pehrson. 1998. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* 393:599–601.

Courties, C., et al. 1994. Smallest eukaryotic organism. *Nature* 370:255.

Fitzgerald, M., T. McKnight, and D. Shippen. 1996. Characterization and developmental patterns of telomerase expression in plants. *Proceedings of the National Academy of Sciences, USA* 93:14422–27.

Gall, J., E. Cohen, and D. Atherton. 1974. The satellite DNAs of *Drosophila virilis. Cold Spring Harbor Symposium on Quantitative Biology* 38:417–21.

Greider, C. 1996. Telomere length regulation. *Annual Review of Biochemistry* 65:337–65.

———. 1999. Telomeres do D-loop-T-loop. *Cell* 97:419–22.

Greider, C., and E. Blackburn. 1996. Telomeres, telomerase and cancer. *Scientific American,* February, 92–97.

Gross, D., and W. Garrard. 1988. Nuclease hypersensitive sites in chromatin. *Annual Review of Biochemistry* 57:159–97.

Grunstein, M. 1992. Histones as regulators of genes. *Scientific American,* October, 68–74B.

Hardison, R. 1999. The evolution of hemoglobin. *American Scientist* 87:126–37.

Horvath, M., et al. 1998. Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell* 95:963–74.

Kavenoff, R., L. Klotz, and B. Zimm. 1974. On the nature of chromosome-sized DNA molecules. *Cold Spring Harbor Symposium on Quantitative Biology* 38:1–8.

Kimura, K., et al. 1999. 13S condensin actively reconfigures DNA by introducing global positive writhe: Implications for chromosome condensation. *Cell* 98:239–48.

Krude, T. 1995. Nucleosome assembly during DNA replication. *Current Biology* 5:1232–34.

Kumar, A., and J. Bennetzen. 1999. Plant retrotransposons. *Annual Review of Genetics* 33:479–532.

Lowary, P., and J. Widom. 1997. Nucleosome packaging and nucleosome positioning of genomic DNA. *Proceedings of the National Academy of Sciences, USA* 94:1183–88.

Luger, K., et al. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389:251–60.

Maniatis, T., et al. 1980. The molecular genetics of human hemoglobins. *Annual Review of Genetics* 14:145–78.

Marcand, S., E. Gilson, and D. Shore. 1997. A protein-counting mechanism for telomere length regulation in yeast. *Science* 275:986–90.

Maxson, R., R. Cohn, and L. Kedes. 1983. Expression and organization of histone genes. *Annual Review of Genetics* 17:239–77.

Melamed, M., T. Lindmo, and M. Mendelsohn, eds. 1990. *Flow Cytometry and Sorting,* 2d ed. New York: Wiley-Liss.

Mitton, J., and M. Grant. 1996. Genetic variation and the natural history of quaking aspen. *BioScience* 46:25–31.

Morse, R. 1993. Nucleosome disruption by transcription factor binding in yeast. *Science* 262:1563–66.

Osheim, Y., and O. Miller, Jr. 1983. Novel amplification and transcriptional activity of chorion genes in *Drosophila melanogaster* follicle cells. *Cell* 33:543–53.

Owen-Hughes, T., et al. 1996. Persistent site-specific remodeling of a nucleosome array by transient action of the SWI/SNF complex. *Science* 273:513–16.

Paranjape, S., R. Kamakaka, and J. Kadonaga. 1994. Role of chromatin structure in the regulation of transcription by RNA polymerase II. *Annual Review of Biochemistry* 63:265–97.

Paulson, J., and U. Laemmli. 1977. The structure of histone depleted metaphase chromosomes. *Cell* 12:817–28.

Pereira, S., et al. 1997. Archaeal nucleosomes. *Proceedings of the National Academy of Sciences, USA* 94:12633–37.

Peterson, C., and J. Workman. 2000. Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Current Opinion in Genetics & Development* 10:187–92.

Petrov, D., et al. 2000. Evidence for DNA loss as a determinant of genome size. *Science* 287:1060–62.

Pruss, D., et al. 1996. An asymmetric model for the nucleosome: A binding site for linker histones inside the DNA gyres. *Science* 274:614–17.

Schulz, H., et al. 1999. Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science* 284:493–95.

Shippen-Lentz, D., and E. Blackburn. 1990. Functional evidence for an RNA template in telomerase. *Science* 247:546–52.

Smit, A., and A. Riggs. 1996. *Tiggers* and other DNA transposon fossils in the human genome. *Proceedings of the National Academy of Sciences, USA* 93:1443–48.

Smith, M., J. Bruhn, and J. Anderson. 1992. The fungus *Armillaria bulbosa* is among the largest and oldest living organisms. *Nature* 356:428–31.

Stamatoyannopoulos, G. 1991. Human hemoglobin switching. *Science* 252:383.

Stark, G., and G. Wahl. 1984. Gene amplification. *Annual Review of Biochemistry* 53:447–91.

Strahl, B., and C. Allis. 2000. The language of covalent histone modifications. *Nature* 403:41–45.

Strunnikov, A. 1998. SMC proteins and chromosome structure. *Trends in Cell Biology* 8:454–59.

Studitsky, V., D. Clark, and G. Felsenfeld. 1994. A histone octamer can step around a transcribing polymerase without leaving the template. *Cell* 76:371–82.

Studitsky, V., et al. 1997. Mechanism of transcription through the nucleosome by eukaryotic RNA polymerase. *Science* 278:1960–63.

Sumner, A. 1990. *Chromosome Banding.* Cambridge, Mass.: Unwin Hyman.

Taylor, J., P. Woods, and W. Hughes. 1957. The organization and duplication of chromosomes as revealed by autoradiographic studies using tritium-labeled thymidine. *Proceedings of the National Academy of Sciences, USA* 43:122–28.

Therman, E., and B. Susman. 1995. *Human Chromosomes,* 3d ed. New York: Springer-Verlag.

Thomas, J. 1999. Histone H1: location and role. *Current Opinion in Cell Biology* 11:312–17.

Tsukiyama, T., and C. Wu. 1997. Chromatin remodeling and transcription. *Current Opinion in Genetics & Development* 7:182–91.

van Steensel, B., and T. de Lange. 1997. Control of telomere length by the human telomeric protein TRF1. *Nature* 385:740–43.

Wade, P., D. Pruss, and A. Wolffe. 1997. Histone acetylation: chromatin in action. *Trends in Biochemical Sciences* 22:128–32.

Wang, Y., and D. Patel. 1993. Solution structure of the human telomeric repeat $d[AG_3(T_2AG_3)_3]$ G-tetraplex. *Structure* 1:263–82.

Whitehouse, I., et al. 1999. Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature* 400:784–87.

Williamson, J. 1994. G-quartet structures in telomeric DNA. *Annual Review of Biophysics and Biomolecular Structure* 23:703–30.

Wolffe, A. 1994. Nucleosome positioning and modification: Chromatin structures that potentiate transcription. *Trends in Biochemical Sciences* 19:240–44.

———. 1995. Histone deviants. *Current Biology* 5:452–54.

Workman, J., and R. Kingston. 1998. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annual Review of Biochemistry* 67:545–79.

Yu, G.-L., et al. 1990. In vivo alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs. *Nature* 344:126–32.

Zhang, X., et al. 1999. Telomere shortening and apoptosis in telomerase-inhibited human tumor cells. *Genes & Development* 13:2388–99.

### Chapter 16 Gene Expression: Control in Eukaryotes

Akira, S., K. Okazaki, and H. Sakano. 1987. Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. *Science* 238:1134–38.

Amedeo, P., et al. 2000. Disruption of the plant gene *MOM* releases transcriptional silencing of methylated genes. *Nature* 405:203–6.

Amig, S., et al. 1994. Transient accumulation of new class II MHC molecules in a novel endocytic compartment in B lymphocytes. *Nature* 369:113–20.

Amit, A., et al. 1986. Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. *Science* 233:747–53.

Arstila, T., et al. 1999. A direct estimate of the human αβ T cell receptor diversity. *Science* 286:958–61.

Autexier, C., and C. Greider. 1996. Telomerase and cancer: revisiting the telomere hypothesis. *Trends in Biochemical Sciences* 21:387–91.

Bashirullah, A., R. Cooperstock, and H. Lipshitz. 1998. RNA localization in development. *Annual Review of Biochemistry* 67:335–94.

Bate, M., and A. Martinez Arias, eds. 1993. *The Development of* Drosophila melanogaster. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Bhattacharyya, M., et al. 1990. The wrinkled-seed character of peas described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* 60:115–22.

Bird, A., and A. Wolffe. 1999. Methylation-induced repression—Belts, braces, and chromatin. *Cell* 99:451–54.

Bischoff, J., et al. 1996. An adenovirus mutant that replicates selectively in p53-deficient human tumor cells. *Science* 274:373–76.

Brown, J., et al. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364:33–39.

Brown, M., and E. Solomon. 1997. Studies on inherited cancers: Outcomes and challenges of 25 years. *Trends in Genetics* 13:202–6.

Burton, D., and J. Moore. 1998. Why do we not have an HIV vaccine and how can we make one? *Nature Medicine Vaccine Supplement* 4:495–98.

Campos-Ortega, J., and V. Hartenstein. 1985. *The Embryonic Development of* Drosophila melanogaster. Berlin: Springer-Verlag.

Carroll, S. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature* 376:479–85.

Casares, F., and R. Mann. 1998. Control of antennal versus leg development in *Drosophila. Nature* 392:723–26.

Caspari, T. 2000. Checkpoints: How to activate p53. *Current Biology* 10:R315–17.

Cavanee, W., and R. White. 1995. The genetic basis of cancer. *Scientific American,* March, 72–79.

Cohen, L. 1987. Diet and cancer. *Scientific American,* November, 42–48.

Cold Spring Harbor Laboratory Press. 1998. *Mechanisms of Transcription.* Cold Spring Harbor Symposia on Quantitative Biology, Volume 63. New York: Cold Spring Harbor.

Collins, E., D. Garboczi, and D. Wiley. 1994. Three-dimensional structure of a peptide extending from one end of a class I MHC site. *Nature* 371:626–29.

Cullen, B., ed. 1993. *Human Retroviruses.* Oxford: IRL Press at Oxford University Press.

Davies, D., E. Padlan, and S. Sheriff. 1990. Antibody-antigen complexes. *Annual Review of Biochemistry* 59:439–73.

Davis, M. 1990. T cell receptor gene diversity and selection. *Annual Review of Biochemistry* 59:475–96.

De Robertis, E., G. Oliver, and C. Wright. 1990. Homeobox genes and the vertebrate body plan. *Scientific American,* July, 46–52.

Doolittle, R. 1995. The multiplicity of domains in proteins. *Annual Review of Biochemistry* 64:287–314.

Driever, W., et al. 1994. Zebrafish: Genetic tools for studying vertebrate development. *Trends in Genetics* 10:152–59.

Duboule, D., ed. 1994. *Guidebook to the Homeobox Genes.* New York: Oxford University Press.

Duke, R., D. Ojcius, and J. Young. 1996. Cell suicide in health and disease. *Scientific American,* December, 80–87.

Dyson, N., et al. 1989. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science* 243:934–37.

Ellenberger, T. 1994. Getting a grip on DNA recognition: Structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. *Current Opinion in Structural Biology* 4:12–21.

Engelhard, V. 1994. How cells process antigens. *Scientific American,* August, 54–61.

Evans, R., and S. Hollenberg. 1988. Zinc fingers: Gilt by association. *Cell* 52:1–3.

Fearon, E. 1997. Human cancer syndromes: Clues to the origin and nature of cancer. *Science* 278:1043–50. (Plus other articles in this section on Frontiers in Cancer Research.)

Federoff, N. 1984. Transposable genetic elements in maize. *Scientific American,* June, 84–98.

Ferré-D'Amaré, A., et al. 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 363:38–45.

Fraser, S., and R. Harland. 2000. The molecular metamorphosis of experimental embryology. *Cell* 100, 41–55.

French, D., R. Laskov, and M. Scharff. 1989. The role of somatic hypermutation in the generation of antibody diversity. *Science* 244:1152–57.

Fugmann, S., et al. 2000. The RAG proteins and V(D)J recombination: Complexes, ends, and transposition. *Annual Review of Immunology* 18:495–527.

Gao, F., et al. 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes. Nature* 397:436–41.

Garcia, K., L. Teyton, and I. Wilson. 1999. Structural basis of T cell recognition. *Annual Review of Immunology* 17:369–97.

Gehring, W. 1999. *Master Control Genes in Development and Evolution: The Homeobox Story.* New Haven, Connecticut: Yale University Press.

Gehring, W., M. Affolter, and T. Bürglin. 1994. Homeodomain proteins. *Annual Review of Biochemistry* 64:487–526.

Gehring, W., et al. 1994. Homeodomain-DNA recognition. *Cell* 78:211–23.

Gierl, A., H. Saedler, and P. Peterson. 1989. Maize transposable elements. *Annual Review of Genetics* 23:71–85.

Goff, S. 1992. Genetics of retroviral integration. *Annual Review of Genetics* 26:527–44.

Goodrich, J., et al. 1997. A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis. Nature* 386:44–51.

Groll, M., et al. 1997. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386:463–71.

Haber, J. 1998. Mating-type gene switching in *Saccharomyces cerevisiae. Annual Review of Genetics* 32:561–99.

Halder, G., P. Callaerts, and W. Gehring. 1995. Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila. Science* 267:1788–92.

Haluska, F., Y. Tsujimoto, and C. Croce. 1987. Oncogene activation by chromosome translocation in human malignancy. *Annual Review of Genetics* 21:321–45.

Hanahan, D., and R. Weinberg. 2000. The hallmarks of cancer. *Cell* 100:57–70.

Harrison, S. 1991. A structural taxonomy of DNA-binding domains. *Nature* 353:715–19.

Hastie, N. 1994. The genetics of Wilm's tumor—A case of disrupted development. *Annual Review of Genetics* 28:523–58.

Heemels, M.-T., and H. Ploegh. 1995. Generation, translocation, and presentation of MHC class I-restricted peptides. *Annual Review of Biochemistry* 64:463–91.

Heinlein, M. 1996. Excision patterns of *Activator (Ac)* and *Dissociation (Ds)* elements in *Zea mays* L.: implications for the regulation of transposition. *Genetics* 144:1851–69.

Hershko, A., and A. Ciechanover. 1998. The ubiquitin system. *Annual Review of Biochemistry* 67:425–79.

Hirao, A., et al. 2000. DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science* 287:1824–27.

Horvitz, H., and J. Sulston. 1990. "Joy of the worm." *Genetics* 126:287–92.

Hunter, T. 2000. Signaling—2000 and beyond. *Cell* 100:113–27.

Johnson, P., and S. McKnight. 1989. Eukaryotic transcriptional regulatory proteins. *Annual Review of Biochemistry* 58:799–839.

Jürgens, G., R. Torres Ruiz, and T. Berleth. 1994. Embryonic pattern formation in flowering plants. *Annual Review of Genetics* 28:351–71.

Karin, M., and T. Hunter. 1995. Transcriptional control by protein phosphorylation: Signal transmission from the cell surface to the nucleus. *Current Biology* 5:747–57.

Katz, R., and A. Skalka. 1994. The retroviral enzymes. *Annual Review of Biochemistry* 63:133–73.

Kennel, S., et al. 1984. Monoclonal antibodies in cancer detection and therapy. *BioScience* 34:150–56.

Kinzler, K., and B. Vogelstein. 1996. Lessons from hereditary colorectal cancer. *Cell* 87:159–70.

Knudson, A., Jr. 1993. Antioncogenes and human cancer. *Proceedings of the National Academy of Sciences, USA* 90:10914–21.

Kussie, P. 1996. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274:948–53.

Laird, P., and R. Jaenisch. 1996. The role of DNA methylation in cancer genetics and epigenetics. *Annual Review of Genetics* 30:441–64.

Landschulz, W., P. Johnson, and S. McKnight. 1988. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* 240:1759–64.

Lawrence, P. 1992. *The Making of a Fly.* Oxford: Blackwell.

Lawrence, P., and G. Morata. 1994. Homeobox genes: Their function in *Drosophila* segmentation and pattern formation. *Cell* 78:181–89.

Lawrence, P., and G. Struhl. 1996. Morphogens, compartments, and pattern: Lessons from *Drosophila? Cell* 85:951–61.

Lebel-Hardenack, S., and S. Grant. 1997. Genetics of sex determination in flowering plants. *Trends in Plant Science* 2:130–36.

Lengauer, C., K. Kinzler, and B. Vogelstein. 1998. Genetic instabilities in human cancers. *Nature* 396:643–49.

Levine, A. 1993. The tumor suppressor genes. *Annual Review of Biochemistry* 62:623–51.

———. 1997. p53, the cellular gatekeeper for growth and division. *Cell* 88:323–31.

Ma, H., and C. dePamphilis. 2000. The ABCs of floral evolution. *Cell* 101:5–8.

Malkin, D. 1994. Germline p53 mutations and heritable cancer. *Annual Review of Genetics* 28:443–65.

Mannervik, M., et al. 1999. Transcriptional coregulators in development. *Science* 284:606–9.

Marx, J. 1984. What do oncogenes do? *Science* 223:673–76.

———. 1994. Oncogenes reach a milestone. *Science* 266:1942–44.

Max, E., J. Seidman, and P. Leder. 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin k constant region gene. *Proceedings of the National Academy of Sciences, USA* 76:3450–54.

McGinnis, W. 1994. A century of homeosis, a decade of homeoboxes. *Genetics* 137:607–11.

McGinnis, W., and R. Krumlauf. 1992. Homeobox genes and axial patterning. *Cell* 68:283–302.

McGinnis, W., and M. Kuziora. 1994. The molecular architects of body design. *Scientific American,* February, 58–66.

McKnight, S. 1991. Molecular zippers in gene regulation. *Scientific American,* April, 54–64.

Melek, M., and M. Gellert. 2000. RAG1/2-mediated resolution of transposition intermediates: two pathways and possible consequences. *Cell* 101:625–33.

Meyerowitz, E. 1994. The genetics of flower development. *Scientific American,* November, 56–65.

MHC Sequencing Consortium. 1999. Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401:921–23 (plus insert).

Milstein, C. 1986. From antibody structure to immunological diversification of immune response. *Science* 231:1261–68.

Mitelman, F., F. Mertens, and B. Johansson. 1997. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nature Genetics Special Issue* 15:417–74.

Ng, H.-H., and A. Bird. 1999. DNA methylation and chromatin modification. *Current Opinion in Genetics & Development* 9:158–63.

Niehrs, C., and N. Pollet. 1999. Synexpression groups in eukaryotes. *Nature* 402:483–87.

Nüsslein-Volhard, C. 1996. Gradients that organize embryo development. *Scientific American,* August, 54–61.

Nüsslein-Volhard, C., and E. Wieschaus. 1980. Mutations affecting segment number and polarity in *Drosophila. Nature* 287:795–801.

O'Brien, S., and M. Dean. 1997. In search of AIDS-resistance genes. *Scientific American,* September, 44–51.

Oettinger, M. 1999. V(D)J recombination: on the cutting edge. *Current Opinion in Cell Biology* 11:325–29.

Oshima, Y., and I. Takano. 1971. Mating types in *Saccharomyces:* Their convertibility and homothallism. *Genetics* 67:327–35.

Pidkowich, M., J. Klenz, and G. Haughn. 1999. The making of a flower: Control of floral meristem identity in *Arabidopsis. Trends in Plant Science* 4:64–70.

Pieters, J. 1997. MHC class II restricted antigen presentation. *Current Opinion in Immunology* 9:89–96.

Rahman, N., and M. Stratton. 1998. The genetics of breast cancer susceptibility. *Annual Review of Genetics* 32:95–121.

Ray, L. 1999. The science of signal transduction. *Science* 284:755–56. (Introduces several articles in a signal transduction review section.)

Reinherz, E., et al. 1999. The crystal structure of a T cell receptor in complex with peptide and MHC class II. *Science* 286:1913–21.

Rhodes, D., and A. Klug. 1993. Zinc fingers. *Scientific American,* February, 56–65.

Riddle, D., et al. 1997. *C. elegans* II: Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Rowen, L., B. Koop, and L. Hood. 1996. The complete 685-kilobase DNA sequence of the human β T cell receptor locus. *Science* 272:1755–62.

Ruddle, F., et al. 1994. Evolution of HOX genes. *Annual Review of Genetics* 28:423–42.

Rudolf, K., et al. 1999. Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* 96:701–12.

Sánchez-Garcia, I. 1997. Consequences of chromosomal abnormalities in tumor development. *Annual Review of Genetics* 31:429–53.

*Scientific American.* 1993. Life, death, and the immune system. September (entire issue, ten articles).

———. 1996. Making headway against cancer. September (entire issue, ten articles).

———. 1998. Defeating AIDS: What will it take? Special Report. July, 81–107 (several articles).

Scott, J., and T. Pawson. 2000. Cell communication: The inside story. *Scientific American,* June, 72–79.

Scott, M. 2000. Development: The natural history of genes. *Cell* 100:27–40.

Smith, G., et al. 1999. Purification and DNA binding properties of the ataxia-telangiectasia gene product ATM. *Proceedings of the National Academy of Sciences, USA* 11134–39.

Stanbridge, E. 1990. Human tumor suppressor genes. *Annual Review of Genetics* 24:615–57.

Stern, L., et al. 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 368:215–21.

St. Johnston, D., and C. Nüsslein-Volhard. 1992. The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 68:201–19.

Strominger, J. 1989. Developmental biology of T-cell receptors. *Science* 244:943–50.

Struhl, K. 1999. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 98:1–4.

Sugden, B. 1993. How some retroviruses got their oncogenes. *Trends in Biochemical Sciences* 18:233–35.

Sulston, J., et al. 1983. The embryonic cell lineage of the nematode, *Caenorhabditis elegans. Developmental Biology* 100:64–119.

Tanaka, H., et al. 2000. A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage. *Nature* 404:42–49.

Tang, H., K. Kuhen, and F. Wong-Staal. 1999. Lentivirus replication and regulation. *Annual Review of Genetics* 33:133–70.

Toyota, M., and J.-P. Issa. 2000. The role of DNA hypermethylation in human neoplasia. *Electrophoresis* 21:329–33.

Travers, A. 1989. DNA conformation and protein binding. *Annual Review of Biochemistry* 58:427–52.

Van Noorden, C., L. Meade-Tollin, and F. Bosman. 1998. Metastasis. *American Scientist* 86:130–41.

Varmus, H. 1988. Retroviruses. *Science* 240:1427–35.

Varmus, H., and N. Nathanson. 1998. Science and the control of AIDS. *Science* 280:1815. (Introducing several articles on AIDS.)

Von Boehmer, H., and P. Kisielow. 1991. How the immune system learns about self. *Scientific American,* October, 74–81.

Waldman, T. 1991. Monoclonal antibodies in diagnosis and therapy. *Science* 252:1657–62.

Weigel, D. 1995. The genetics of flower development: from floral induction to ovule morphogenesis. *Annual Review of Genetics* 29:19–39.

Weigel, D., and O. Nilsson. 1995. A developmental switch sufficient for flower initiation in diverse plants. *Nature* 377:495–500.

Weissman, B., et al. 1987. Introduction of a normal human chromosome 11 into a Wilm's tumor cell line controls its tumorigenic expression. *Science* 236:175–80.

Zhu, T., et al. 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* 391:594–97.

Ziegler, A., et al. 1994. Sunburn and p53 in the onset of skin cancer. *Nature* 372:773–76.

### Chapter 17 Non-Mendelian Inheritance

Armstrong, M., V. Blok, and M. Phillips. 2000. A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida. Genetics* 154:181–92.

Avise, J. 1991. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annual Review of Genetics* 25:45–69.

Birger, Y., et al. 1999. The imprinting box of the mouse *Igf2r* gene. *Nature* 397:84–88.

Blackburn, E., and K. Karrer. 1986. Genomic reorganization in ciliated protozoans. *Annual Review of Genetics* 20:501–21.

Brannan, C., and M. Bartolomei. 1999. Mechanisms of genomic imprinting. *Current Opinion in Genetics & Development* 9:164–70.

Buiting, K., et al. 1995. Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nature Genetics* 9:395–400.

Chang, D., and D. Clayton. 1987. A mammalian mitochondrial RNA processing activity contains nucleus-encoded RNA. *Science* 235:1178–84.

Christian, J., and C. Lemunyan. 1958. Adverse effects of crowding on lactation and reproduction of mice and two generations of their progeny. *Endocrinology* 63:517–29.

Ciferri, O., and L. Dure, III, eds. 1983. *Structure and Function of Plant Genomes.* New York: Plenum Press.

Clewell, D., ed. 1993. *Bacterial Conjugation.* New York: Plenum Press.

Dilts, J., and R. Quackenbush. 1986. A mutation in the R body-coding sequence destroys expression of the killer trait in *P. tetraurelia. Science* 232:641–43.

Evans, R., K. Oakley, and G. Clark-Walker. 1985. Elevated levels of petite formation in strains of *Saccharomyces cerevisiae* restored to respiratory competence. I: Association of both high and moderate frequencies of petite mutant formation with the presence of aberrant mitochondrial DNA. *Genetics* 111:389–402.

Gyllensten, U., et al. 1991. Paternal inheritance of mitochondrial DNA in mice. *Nature* 352:255–57.

Hardy, K. 1994. *Plasmids.* Oxford: Oxford University Press.

Hoeh, W., K. Blakley, and W. Brown. 1991. Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA. *Science* 251:1488–90.

Hurt, E., and G. Schatz. 1987. A cytosolic protein contains a cryptic mitochondrial targeting signal. *Nature* 325:499–503.

Ippen-Ihler, K., and E. Minkley, Jr. 1986. The conjugation system of F, the fertility factor of *Escherichia coli. Annual Review of Genetics* 20:593–624.

Johns, D. 1996. The other human genome: mitochondrial DNA and disease. *Nature Medicine* 2:1065–68.

Kiberstis, P., ed. 1999. Mitochondria make a comeback. *Science* 283:1475–97. (Four articles on evolution, diseases, function, and inheritance.)

Kuroiwa, T., and H. Uchida. 1996. Organelle division and cytoplasmic inheritance. *BioScience* 46:827–35.

Lalande, M. 1996. Parental imprinting and human disease. *Annual Review of Genetics* 30:173–95.

Landis, W. 1987. Factors determining the frequency of the killer trait within populations of the *Paramecium aurelia* complex. *Genetics* 115:197–205.

Lang, B., M. Gray, and G. Burger. 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annual Review of Genetics* 33:351–97.

Lang, B., et al. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* 387:493–97.

Larsson, N., and D. Clayton. 1995. Molecular genetic aspects of human mitochondrial disorders. *Annual Review of Genetics* 29:151–78.

Lewin, B. 1998. The mystique of epigenetics. *Cell* 93:301–3. (Followed by seven articles on epigenetic inheritance.)

Lightowlers, R., et al. 1997. Mammalian mitochondrial genetics: Heredity, heteroplasmy and disease. *Trends in Genetics* 13:450–55.

Luft, R. 1994. The development of mitochondrial medicine. *Proceedings of the National Academy of Sciences, USA* 91:8731–38.

Lunt, D., and B. Hyman. 1997. Animal mitochondrial DNA recombination. *Nature* 387:247.

Mannella, C., M. Marko, and K. Buttle. 1997. Reconsidering mitochondrial structure: New views of an old organelle. *Trends in Biochemical Sciences* 22:37–38.

Margulis, L. 1995. *Symbiosis in Cell Evolution,* 2d ed. San Francisco: Freeman.

McKusick, V. 1998. *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders,* 12th ed. Baltimore: Johns Hopkins University Press.

Neale, D., K. Marshall, and R. Sederoff. 1989. Chloroplast and mitochondrial DNA are paternally inherited in *Sequoia sempervirens* D. Don Endl. *Proceedings of the National Academy of Sciences, USA* 86:9347–49.

Perrimon, N., L. Engstrom, and A. Mahowald. 1989. Zygotic lethals with specific maternal effect phenotypes in *Drosophila melanogaster.* I. Loci on the X chromosome. *Genetics* 121:333–52.

Pfanner, N., and W. Neupert. 1990. The mitochondrial protein import apparatus. *Annual Review of Biochemistry* 59:331–53.

Preer, J., Jr. 1997. Whatever happened to *Paramecium* genetics? *Genetics* 145:217–25.

Preer, J., Jr., L. Preer, and A. Jurand. 1974. Kappa and other endosymbionts in *Paramecium aurelia. Bacteriological Reviews* 38:113–63.

Razin, A., and H. Cedar. 1994. DNA methylation and genomic imprinting. *Cell* 77:473–76.

Rhoades, M. 1946. Plastid mutations. *Cold Spring Harbor Symposium on Quantitative Biology* 11:202–7.

Rodgers, J. 1991. Mechanisms Mendel never knew. *Mosaic* 22:2–11. (Entire issue on nontraditional inheritance.)

Sager, R. 1972. *Cytoplasmic Genes and Organelles.* New York: Academic Press.

Saville, B., Y. Kohli, and J. Anderson. 1998. mtDNA recombination in a natural population. *Proceedings of the National Academy of Sciences, USA* 95:1331–35.

Shadel, G., and D. Clayton. 1997. Mitochondrial DNA maintenance in vertebrates. *Annual Review of Biochemistry* 66:409–35.

Sonneborn, T. 1950. Methods in the general biology and genetics of Paramecium aurelia. *Journal of Experimental Zoology* 113:87–148.

Stephanou, G., and S. Alahiotis. 1983. Non-Mendelian inheritance of “heat-sensitivity” in *Drosophila melanogaster. Genetics* 103:93–107.

Stoneking, M., and H. Soodyall. 1996. Human evolution and the mitochondrial genome. *Current Opinion in Genetics and Development* 6:731–36.

Summers, D. 1996. *The Biology of Plasmids.* Oxford: Blackwell Scientific.

Thomas, C., ed. 1989. *Promiscuous Plasmids of Gram-Negative Bacteria.* London: Academic Press.

Tilghman, S. 1999. The sins of the fathers and mothers: Genomic imprinting in mammalian development. *Cell* 96:185–93.

Unseld, M., et al. 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics* 15:57–61.

Walbot, V., and E. Coe, Jr. 1979. Nuclear gene *iojap* conditions a programmed change to ribosome-less plastids in *Zea mays. Proceedings of the National Academy of Sciences, USA* 76:2760–64.

Wallace, D. 1992. Diseases of the mitochondrial DNA. *Annual Review of Biochemistry* 61:1175–212.

———. 1997. Mitochondrial DNA in aging and disease. *Scientific American,* August, 40–47.

Wallace, D., et al. 1988. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science* 242:1427–30.

Warren, G., and W. Wickner. 1996. Organelle inheritance. *Cell* 84:395–400.

Wickner, R., 1986. Double-stranded RNA replication in yeast: The killer system. *Annual Review of Biochemistry* 55:373–95.

Wiener, M., et al. 1997. Crystal structure of colicin Ia. *Nature* 385:461–64.

Williamson, D., and D. Poulson. 1979. Sex ratio organisms (spiroplasmas) of *Drosophila. The Mycoplasmas* 3:175–208.

Yaffe, M., et al. 1996. Microtubules mediate mitochondrial distribution in fission yeast. *Proceedings of the National Academy of Sciences, USA* 93:11664–68.

### Chapter 18 Quantitative Inheritance

American Association for the Advancement of Science. 1994. Genes and behavior. *Science* 264:1685–739.

Bennett, S., and J. Todd. 1996. Human type 1 diabetes and the insulin gene: Principles of mapping polygenes. *Annual Review of Genetics* 30:343–70.

Bouchard, T., Jr., and M. McGue. 1981. Familial studies of intelligence: A review. *Science* 212:1055–59.

Carson, R., and M. Rothstein, eds. 1999. *Behavioral Genetics: The Clash of Culture and Biology.* Baltimore: Johns Hopkins University Press.

Cheverud, J., and E. Routman. 1995. Epistasis and its contribution to genetic variance components. *Genetics* 139:1455–61.

Cheverud, J., et al. 1996. Quantitative trait loci for murine growth. *Genetics* 142:1305–19.

Crow, J. 1957. Genetics of insect resistance to chemicals. *Annual Review of Entomology* 2:227–46.

———. 1997. Birth defects, Jimson weed and bell curves. *Genetics* 147:1–6.

Daniels, S., et al. 1996. A genome-wide search for quantitative trait loci underlying asthma. *Nature* 383:247–50.

Devlin, B., M. Daniels, and K. Roeder. 1997. The heritability of IQ. *Nature* 388:468–71.

de Waal, F. 1999. The end of nature versus nurture. *Scientific American,* December, 94–99.

Eshed, Y., and D. Zamir. 1996. Less-than-additive epistatic interactions of quantitative trait loci in tomato. *Genetics* 143:1807–17.

Falconer, D. 1992. Early selection experiments. *Annual Review of Genetics* 26:1–14.

———. 1995. *Introduction to Quantitative Genetics,* 4th ed. London: Longman.

Greenspan, R. 1995. Understanding the genetic construction of behavior. *Scientific American,* April, 72–78.

Hall, J. 1996. Twinning: Mechanisms and genetic implications. *Current Opinion in Genetics and Development* 6:343–47.

Harrison, G., and J. Owen. 1964. Studies on the inheritance of human skin color. *Annals of Human Genetics* 28:27–37.

Holt, S. 1961. Inheritance of dermal ridge patterns. In *Recent Advances in Human Genetics,* edited by L. Penrose, pp. 101–19. London: J. and A. Churchill.

Horgan, J. 1993. Eugenics revisited. *Scientific American,* June, 122–31.

Keightley, P., et al. 1996. A genetic map of quantitative trait loci for body weight in the mouse. *Genetics* 142:227–35.

Laxova, R. 1998. Lionel Sharples Penrose, 1898–1972: A personal memoir in celebration of the centenary of his birth. *Genetics* 150:1333–40.

Li, A., et al. 1997. Epistasis for three grain yield components in rice (*Oryza sativa L.*). *Genetics* 145:453–65.

Lynch, M., and B. Walsh. 1997. *Genetics and Analysis of Quantitative Traits.* Sunderland, Mass.: Sinauer Associates.

Mackintosh, N., ed. 1995. *Cyril Burt: Fraud or Framed?* Oxford: Oxford University Press.

Paterson, A., et al. 1991. Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. *Genetics* 127:181–97.

Plomin, R. 1990. The role of inheritance in behavior. *Science* 248:183–88.

———. 1999. Genetics and general cognitive ability. *Nature* 402:c25–c29.

Plomin, R., and J. DeFries. 1998. The genetics of cognitive abilities and disabilities. *Scientific American,* May, 62–69.

Powledge, T. 1993. The genetic fabric of human behavior. *BioScience* 43:362–67.

———. 1993. The inheritance of behavior in twins. *BioScience* 43:420–24.

Rose, R., et al. 1990. Social contact and sibling similarity: Facts, issues, and red herrings. *Behavioral Genetics* 20:763–78.

Tanksley, S. 1993. Mapping polygenes. *Annual Review of Genetics* 27:205–33.

Velmalal, R., et al. 1999. A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle. *Animal Genetics* 30:136–43.

Vieira, C., et al. 2000. Genotype-environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster. Genetics* 154:213–27.

Woolf, C. 1990. Multifactorial inheritance of common white markings in the Arabian horse. *Journal of Heredity* 81:250–56.

Yashin, A., I. Iachine, and J. Harris. 1999. Half of the variation in susceptibility to mortality is genetic: Findings from Swedish twin survival data. *Behavior Genetics* 29:11–19.

### Chapter 19 Population Genetics: The Hardy-Weinberg Equilibrium and Mating Systems

Bittles, A., and J. Neel. 1994. The costs of human inbreeding and their implications for variations at the DNA level. *Nature Genetics* 8:117–21.

Bittles, A., et al. 1991. Reproductive behavior and health in consanguineous marriages. *Science* 252:789–94.

Buss, D. 1985. Human mate selection. *American Scientist* 73:47–51.

Cavalli-Sforza, L., P. Menozzi, and A. Piazza. 1996. *The History and Geography of Human Genes.* Princeton, N.J.: Princeton University Press.

Crow, J. 1987. Population genetics history: A personal view. *Annual Review of Genetics* 21:1–22.

———. 1988. Eighty years ago: The beginnings of population genetics. *Genetics* 119:473–76.

———. 1988. Sewall Wright (1889–1988). *Genetics* 119:1–4.

De Finetti, B. 1926. Considerazioni matematiche sul l'ereditarieta mendeliana. *Metron* 6:1–41.

Dudash, M., and D. Carr. 1998. Genetics underlying inbreeding depression in *Mimulus* with contrasting mating systems. *Nature* 393:682–84.

Fisher, R. 1930. *The Genetical Theory of Natural Selection.* Oxford: Clarendon Press.

Frankham, R. 1995. Conservation genetics. *Annual Review of Genetics* 29:305–27.

Haldane, J. 1932. *The Causes of Evolution.* New York: Harper & Row.

Hardy, G. 1908. Mendelian proportions in a mixed population. *Science* 28:49–50.

Hill, W. 1995. Sewall Wright's "systems of mating." *Genetics* 143:1499–1506.

Jimenez, J., et al. 1994. An experimental study of inbreeding depression in a natural habitat. *Science* 266:271–73.

Kimura, M. 1988. Thirty years of population genetics with Dr. Crow. *Japanese Journal of Genetics* 63:1–10.

May, R. 1995. The cheetah controversy. *Nature* 374:309–10.

Maynard Smith, J. 1989. *Evolutionary Genetics.* Oxford: Oxford University Press.

Ralls, K., J. Ballou, and A. Templeton. 1988. Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conservation Biology* 2:185–93.

Russell, E. 1989. Sewall Wright's contributions to physiological genetics and to inbreeding theory and practice. *Annual Review of Genetics* 23:1–18.

Saccheri, I., et al. 1998. Inbreeding and extinction in a butterfly metapopulation. *Nature* 392:491–94.

Sarkar, S. 1992. A centenary reassessment of J.B.S. Haldane, 1892–1964. *BioScience* 42:777–85.

Spiess, E. 1989. *Genes in Populations,* 2d ed. New York: Wiley-Liss.

Weinberg, W. 1908. Über den Nachweis der Vererbung beim Menschen. *Jahreshefte Verein fuer Vaterlaendische Naturkunde in Wuerttemberg* 64:368–82.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.

———. 1968–78. *Evolution and the Genetics of Populations: A Treatise.* 4 vols. Chicago: University of Chicago Press.

### Chapter 20 Population Genetics: Processes That Change Allelic Frequencies

Cameron, R. 1992. Change and stability in *Cepaea* populations over 25 years: A case of climatic selection. *Proceedings of the Royal Society of London, series B* 248:181–87.

Caro, T., and M. Laurenson. 1994. Ecological and genetic factors in conservation: A cautionary tale. *Science* 263:485–86.

Carson, H. 1997. Sexual selection: A driver of genetic change in Hawaiian *Drosophila. The Journal of Heredity* 88:343–52.

Castle, W. 1903. The laws of heredity of Galton and Mendel, and some laws governing race improvement by selection. *Proceedings of the American Academy of Arts and Sciences* 39:223–42.

Crow, J. 1987. Population genetics history: A personal view. *Annual Review of Genetics* 21:1–22.

———. 1990. Fisher's contributions to genetics and evolution. *Theoretical Population Biology* 38:263–75.

———. 1992. Centennial: J.B.S. Haldane, 1892–1964. *Genetics* 130:1–6.

———. 1992. Erwin Schrödinger and the hornless cattle problem. *Genetics* 130:237–39.

Diamond, J. 1988. Founding fathers and mothers. *Natural History,* June, 10–15.

———. 1994. Pitcairn before the Bounty. *Nature* 369:608–9.

Dobzhansky, T. 1958. Genetics of natural populations. XXVII. The genetic changes in populations of *Drosophila pseudoobscura* in the American Southwest. *Evolution* 12:385–401.

Ganetzky, B. 1999. Yuichiro Hiraizumi and forty years of segregation distortion. *Genetics* 152:1–4.

Gibbons, A. 1994. Genes point to a new identity for Pacific pioneers. *Science* 263:32–33.

Hill, G. 1991. Plumage coloration is a sexually selected indicator of male quality. *Nature* 350:337–39.

Houle, D., et al. 1992. The genomic mutation rate for fitness in *Drosophila. Nature* 359:58–60.

Kimura, M. 1955. Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences, USA* 41:144–50.

———. 1988. Thirty years of population genetics with Dr. Crow. *Japanese Journal of Genetics* 63:1–10.

Koenig, W. 1989. Sex-biased dispersal in the contemporary United States. *Ethology and Sociobiology* 10:263–78.

Koenig, W., S. Albano, and J. Dickson. 1991. A comparison of methods to partition selection acting via components of fitness: Do larger male bullfrogs have greater hatching success? *Journal of Evolutionary Biology* 4:309–20.

Lewontin, R. 1985. Population genetics. *Annual Review of Genetics* 19:81–102.

Lummaa, V., et al. 1998. Natural selection on human twinning. *Nature* 394:533–34.

Lyttle, T. 1991. Segregation distorters. *Annual Review of Genetics* 25:511–57.

Mallett, J., et al. 1990. Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics* 124:921–36.

Maynard Smith, J. 1991. The population genetics of bacteria. *Proceedings of the Royal Society of London, series B* 245:37–41.

McCauley, D. 1991. Genetic consequences of local population extinction and recolonization. *Trends in Ecology and Evolution* 6:5–8.

Menotti-Raymond, M., and S. O'Brien. 1993. Dating the genetic bottleneck of the African cheetah. *Proceedings of the National Academy of Sciences, USA* 90:3172–76.

Pomiankowski, A., and L. Hurst. 1993. Siberian mice upset Mendel. *Nature* 363:396–97.

Powers, D., et al. 1991. Genetic mechanisms for adapting to a changing environment. *Annual Review of Genetics* 25:629–59.

Pyke, G. 1991. What does it cost a plant to produce floral nectar? *Nature* 350:58–59.

Raymond, M., et al. 1996. Frequency-dependent maintenance of left handedness in humans. *Proceedings of the Royal Society of London, series B* 263:1627–33.

Reznick, D., et al. 1997. Evaluation of the rate of evolution in natural populations of guppies *(Poecilia reticulata). Science* 275:1934–37.

Rood, S., et al. 1988. Gibberellins: A phytohormonal basis for heterosis in maize. *Science* 241:1216–18.

Schaeffer, S., and E. Miller. 1992. Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. *Genetics* 132:471–80.

Smith, T. 1993. Disruptive selection and the genetic basis of bill size polymorphism in the African finch *Pyrenestes. Nature* 363:618–20.

Spiess, E. 1989. *Genes in Populations,* 2d ed. New York: Wiley-Liss.

Tamarin, R., and T. Kunz. 1974. *Microtus breweri. Mammalian Species* 45:1–3.

Thoday, J., and J. Gibson. 1962. Isolation by disruptive selection. *Nature* 193:1164–66.

Wallace, B. 1991. *Fifty Years of Genetic Load. An Odyssey.* Ithaca, N.Y.: Cornell University Press.

Weber, K. 1992. How small are the smallest selectable domains of form? *Genetics* 130:345–53.

Williams, G. 1992. *Natural Selection: Domains, Levels, and Challenges.* New York: Oxford University Press.

### Chapter 21 Genetics of the Evolutionary Process

Ayala, F. 1997. Vagaries of the molecular clock. *Proceedings of the National Academy of Sciences, USA* 94:7776–83.

Barbadilla, A., L. King, and R. Lewontin. 1996. What does electrophoretic variation tell us about protein variation? *Molecular Biology and Evolution* 13:427–32.

Barlow, G. 1991. Nature-nurture and the debates surrounding ethology and sociobiology. *American Zoologist* 31:286–96.

Bethell, T. 1976. Darwin's mistake. *Harper's,* February, 70–75.

Bowcock, A., et al. 1991. Drift, admixture, and selection in human evolution: A study with DNA polymorphisms. *Proceedings of the National Academy of Sciences, USA* 88:839–43.

Brakefield, P. 1987. Industrial melanism: Do we have all the answers? *Trends in Ecology and Evolution* 2:117–22.

Britton-Davidian, J., et al. 2000. Rapid chromosomal evolution in island mice. *Nature* 403:158.

Bush, G. 1975. Modes of animal speciation. *Annual Review of Ecology and Systematics* 6:339–64.

Buss, D. 1994. The strategies of human mating. *American Scientist* 82:238–49.

Carson, H. 1987. The genetic system, the deme, and the origin of species. *Annual Review of Genetics* 21:405–23.

Cavalli-Sforza, L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes.* Princeton, N.J.: Princeton University Press.

Cohn, J. 1992. Naked mole-rats. *BioScience* 42:86–89.

Conover, D., and D. Van Voorhees. 1990. Evolution of a balanced sex ratio by frequency-dependent selection in a fish. *Science* 250:1556–58.

Coppens, Y. 1994. East side story: The origin of humankind. *Scientific American,* May, 88–95.

Coyne, J. 1992. Genetics and speciation. *Nature* 355:511–15.

Crow, J. 1987. Population genetics history: A personal view. *Annual Review of Genetics* 21:1–22.

Darwin, C. 1845. *The Voyage of the Beagle.* 1962 ed. Garden City, N.Y.: Doubleday.

———. 1859. *The Origin of Species by Means of Natural Selection or the Preservation of Favored Races in the Struggle for Life.* London: John Murray.

Eldredge, N., and S. Gould. 1972. Punctuated equilibria: An alternative to phyletic gradualism. *In Models in Paleobiology,* edited by T. Schopf, pp. 85–115. San Francisco: Freeman.

Elena, S., V. Cooper, and R. Lenski. 1996. Punctuated evolution caused by selection of rare beneficial mutations. *Science* 272:1802–4.

Emlen, S., P. Wrege, and N. Demong. 1995. Making decisions in the family: An evolutionary perspective. *American Scientist* 83:148–57.

Ford, E. 1979. *Ecological Genetics,* 4th ed. London: Chapman and Hall.

Futuyma, D. 1997. *Evolutionary Biology,* 3d ed. Sunderland, Mass.: Sinauer Associates.

Gibbons, A. 1996. On the many origins of species. *Science* 273:1496–99.

———. 1998. Calibrating the mitochondrial clock. *Science* 279:28–29.

Gillespie, J. 1994. *The Causes of Molecular Evolution.* Oxford: Oxford University Press. (Reprint)

Gojobori, T., E. Moriyama, and M. Kimura. 1990. Molecular clock of viral evolution, and the neutral theory. *Proceedings of the National Academy of Sciences, USA* 87:10015–18.

Gould, S. 1994. The evolution of life on earth. *Scientific American,* October, 84–91.

Gould, S., and N. Eldredge. 1993. Punctuated equilibrium comes of age. *Nature* 366:223–27.

Grant, B., and P. Grant. 1993. Evolution of Darwin's finches caused by a rare climatic event. *Proceedings of the Royal Society of London, series B* 251:111–17.

Grant, B., et al. 1998. Geographic and temporal variation in the incidence of melanism in peppered moth populations in America and Britain. *The Journal of Heredity* 89:465–71.

Grant, P. 1999. *Ecology and Evolution of Darwin's Finches.* Princeton, N.J.: Princeton University Press.

Hamilton, W. 1964. The genetical evolution of social behavior, I, II: *Journal of Theoretical Biology* 7:1–52.

Hanson, B., et al., eds. 1999. The diversity of evolution. *Science* 284:2105–48. (Nine articles and viewpoints.)

Hays, B. 1999. Experimental Lamarckism. *American Scientist* 87:494–98.

Hutter, P., J. Roote, and M. Ashburner. 1990. A genetic basis for the inviability of hybrids between sibling species of *Drosophila. Genetics* 124:909–20.

Johnson, T., et al. 1996. Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science* 273:1091–93.

Kimura, M. 1985. *The Neutral Theory of Molecular Evolution.* New York: Cambridge University Press.

———. 1987. Molecular evolutionary clock and the neutral theory. *Journal of Molecular Evolution* 26:24–33.

Koehn, R. 1970. Functional and evolutionary dynamics of polymorphic esterases in catostomid fishes. *American Fisheries Society, Transactions* 99:219–28.

Kuhn, T. 1996. *The Structure of Scientific Revolutions,* 3d ed. Chicago: University of Chicago Press.

Lewin, R. 1988. Molecular clocks turn a quarter century. *Science* 239:561–63.

Lewontin, R. 1985. Population genetics. *Annual Review of Genetics* 19:81–102.

———. 1991. Twenty-five years ago in GENETICS: Electrophoresis in the development of evolutionary genetics: Milestone or millstone? *Genetics* 128:657–62.

Lewontin, R., and J. Hubby. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II: *Genetics* 54:595–609.

Lindström, L., R. Alatalo, and J. Mappes. 1997. Imperfect Batesian mimicry—The effects of the frequency and the distastefulness of the model. *Proceedings of the Royal Society of London, Series B* 264:149–53.

Losey, J., et al. 1997. A polymorphism maintained by opposite patterns of parasitism and predation. *Nature* 388:269–72.

Losos, J., K. Warheit, and T. Schoener. 1997. Adaptive differentiation following experimental island colonization in *Anolis* lizards. *Nature* 387:70–73.

Lynch, M., and P. Jarrell. 1993. A method for calibrating molecular clocks and its application to animal mitochondrial DNA. *Genetics* 135:1197–208.

May, R. 1992. How many species inhabit the earth? *Scientific American,* October, 42–48.

Mayr, E. 1997. The objects of selection. *Proceedings of the National Academy of Sciences, USA* 94:2091–94.

Mogie, M. 1996. Malthus and Darwin: World views apart. *Evolution* 50:2086–88.

National Academy of Sciences. 1999. *Science and Creationism—A View from the National Academy of Sciences,* 2d ed. Washington: National Academy Press.

Niemalä, P., and J. Tuomi. 1987. Does the leaf morphology of some plants mimic caterpillar damage? *Oikos* 50:256–57.

Nijhout, H. 1994. Developmental perspectives on evolution of butterfly mimicry. *BioScience* 44:148–57.

Nowak, M., R. May, and K. Sigmund. 1995. The arithmetics of mutual help. *Scientific American,* June, 76–81.

Packer, C., et al. 1991. A molecular genetic analysis of kinship and cooperation in African lions. *Nature* 351:562–65.

Pfennig, D., and P. Sherman. 1995. Kin recognition. *Scientific American,* June, 98–103.

Ritland, D., and L. Brower. 1991. The viceroy butterfly is not a Batesian mimic. *Nature* 350:497–98.

Sato, A., et al. 1999. Phylogeny of Darwin's finches as revealed by mtDNA sequences. *Proceedings of the National Academy of Sciences, USA* 96:5101–6.

Scherer, S. 1990. The protein molecular clock: Time for a reevaluation. *Evolutionary Biology* 24:83–106.

Schliewen, U., D. Tautz, and S. Pääbo. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* 368:629–32.

Schmidt, K. 1996. Creationists evolve new strategy. *Science* 273:420–22.

Sherman, P., J. Jarvis, and S. Braude. 1992. Naked mole rats. *Scientific American,* August, 72–78.

Skibinski, D., M. Woodwark, and R. Ward. 1993. A quantitative test of the neutral theory using pooled allozyme data. *Genetics* 135:233–48.

Slatkin, M., ed. 1995. *Exploring Evolutionary Biology.* Sunderland, Mass.: Sinauer Associates.

Stiassny, M., and A. Meyer. 1999. Cichlids of the rift lakes. *Scientific American,* February, 64–69.

Tamarin, R., and M. Sheridan. 1987. Behavior-genetic mechanisms of population regulation in microtine rodents. *American Zoologist* 27:921–27.

Thompson, K. 1998. 1798: Darwin and Malthus. *American Scientist* 86:226–29.

Tregenza, T., and R. Butlin. 1999. Speciation without isolation. *Nature* 400:311–12. (With two accompanying research articles.)

Vincek, V., et al. 1997. How large was the founding population of Darwin's finches? *Proceedings of the Royal Society of London, series B* 264:111–18.

Williams, G. 1992. *Natural Selection: Domains, Levels, and Challenges.* New York: Oxford University Press.

Wilson, E. 1975. *Sociobiology: The New Synthesis.* Cambridge, Mass.: Harvard University Press.

Young, W. 1985. *Fallacies of Creationism.* Calgary, Alberta: Detselig Enterprises, Ltd.

Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins,* edited by V. Bryson and H. Vogel, pp. 97–166. New York: Academic Press.

# GLOSSARY

**acentric fragment** A chromosomal piece without a centromere.

**acrocentric chromosome** A chromosome whose centromere lies very near one end.

**acron** The anterior end of the arthropod embryo from which eyes and antennae develop.

**activation energy** *See* free energy of activation ($\Delta G^{\ddagger}$).

**activator** A eukaryotic specific transcription factor that binds to an enhancer, often far upstream of a promoter.

**active site** The part of an enzyme where the actual enzymatic function is performed.

**adaptive mutation** *See* directed mutation.

**adaptive value** *See* fitness.

**additive model** A mechanism of quantitative inheritance in which alleles at different loci either add an amount to the phenotype or add nothing.

**adenine** *See* purines.

**adjacent-1 segregation** A separation of centromeres during meiosis in a reciprocal translocation heterozygote so that unbalanced zygotes are produced.

**adjacent-2 segregation** Separation of centromeres during meiosis in a translocation heterozygote so that homologous centromeres are pulled to the same pole.

**A DNA** The form of DNA with high water content; it has tilted base pairs and more base pairs per turn than does B DNA.

**affected** Individuals in a pedigree that exhibit the specific phenotype under study.

**allele** Alternative form of a gene.

**allelic exclusion** A process whereby only one immunoglobulin light chain and one heavy chain gene are transcribed in any one cell; the other genes are repressed.

**allopatric speciation** Speciation in which the evolution of reproductive isolating mechanisms occurs during the physical separation of the populations.

**allopolyploidy** Polyploidy produced by the hybridization of two species.

**allosteric protein** A protein whose shape is changed when it binds a particular molecule. In the new shape, the protein's ability to react to a second molecule is altered.

**allotype** Mutant of a nonvariant part of an immunoglobulin gene that follows the rules of simple Mendelian inheritance.

**allozygosity** Homozygosity in which the two alleles are alike but unrelated. *See* autozygosity.

**allozymes** Forms of an enzyme, controlled by alleles of the same locus, that differ in electrophoretic mobility. *See* isozymes.

**alternate segregation** A separation of centromeres during meiosis in a reciprocal translocation heterozygote so that balanced gametes are produced.

**alternative splicing** Various ways of splicing out introns in eukaryotic premessenger RNAs so that one gene produces several different messenger RNA and protein products.

**altruism** A form of behavior in which an individual risks lowering its fitness for the benefit of another.

**Alu family** A dispersed, intermediately repetitive DNA sequence found in the human genome about 300,000 times. The sequence is about 300 bp long. The name *Alu* comes from the restriction endonuclease that cleaves it.

**aminoacyl-tRNA synthetases** Enzymes that attach amino acids to their proper transfer RNAs.

**amphidiploid** An organism produced by hybridization of two species, followed by somatic doubling. It is an allotetraploid that appears to be a normal diploid.

**anagenesis** The evolutionary process whereby one species evolves into another without any splitting of the phylogenetic tree. *See* cladogenesis.

**anaphase** The stage of mitosis and meiosis in which sister chromatids or homologous chromosomes are separated by spindle fibers.

**anaphase A** The stage of anaphase in which chromatids are separated by the shortening of kinetochore microtubules.

**anaphase B** The stage of anaphase in which chromatids are separated by the general elongation of the spindle.

**anaphase-promoting complex (APC)** Protein complex that breaks down cyclin B and promotes anaphase among its various roles in controlling the cell cycle. (Also called the cyclosome.)

**aneuploids** Individuals or cells exhibiting aneuploidy.

**aneuploidy** The condition of a cell or of an organism that has additions or deletions of whole chromosomes.

**angiosperms** Plants whose seeds are enclosed within an ovary. Flowering plants.

**antibody** A protein produced by a B lymphocyte that protects the organism against antigens.

**anticoding strand** The DNA strand that forms the template for both the transcribed messenger RNA and the coding strand.

**anticodon** The three-base sequence on transfer RNA complementary to a codon on messenger RNA.

**antigen** A foreign substance capable of triggering an immune response in an organism.

**antimutator mutations** Mutations of DNA polymerase that decrease the overall mutation rate of a cell or of an organism.

**anti-oncogene** A gene that represses malignant growth and whose absence results in malignancy (e.g., *retinoblastoma*).

**antiparallel strands** Strands, as in DNA, that run in opposite directions with respect to their $3'$ and $5'$ends.

**antisense RNA** RNA product of *mic* (*m*RNA-*i*nterfering *c*omplementary RNA) genes that regulates another gene by base pairing with, and thus blocking, its messenger RNA.

**antisense strand** *See* anticoding strand.

**anti-sigma factor** A protein that interferes with the action of a sigma factor.

**antiterminator protein** A protein that, when bound at its normal attachment sites, lets RNA polymerase read through normal terminator sequences (e.g., the N- and Q-gene products of phage $\lambda$).

**AP endonucleases** Endonucleases that initiate excision repair at apurinic and apyrimidinic sites on DNA.

**apoptosis** Programmed cell death.

**archaea** Highly specialized, bacterialike organisms that make up the third kingdom of life on earth along with the bacteria and the eukaryotes. Identified by Carl Woese in 1977 based on ribosomal RNA sequences. Most are thermophilic, halophilic, or methanogenic.

**ascospores** Haploid spores found in the asci of Ascomycete fungi.

**ascus** The sac in Ascomycete fungi that holds the ascospores.

**A (aminoacyl) site** The site on the ribosome occupied by an aminoacyl-tRNA just prior to peptide bond formation.

**assignment test** A test that determines whether a locus is on a specific chromosome by observing the concordance of the locus and the specific chromosome in hybrid cell lines.

**assortative mating** The mating of individuals with similar phenotypes.

**aster** Configuration at the centrosome with microtubules radiating out in all directions.

**ataxia-telangiectasia** A disease in human beings caused by a defect in X-ray-induced repair mechanisms.

**attenuator region** A control region at the promoter end of repressible amino acid operons that exerts transcriptional control based on the translation of a small leader peptide gene.

**attenuator stem** *See* terminator stem.

**autogamy** Nuclear reorganization in a single *Paramecium* cell similar to the changes that occur during conjugation.

**autonomously replicating sequence (ARS)** Eukaryotic site of the initiation of DNA replication consisting of an 11 base-pair consensus sequence and several other sequences covering 100–200 base pairs.

**autopolyploidy** Polyploidy in which all the chromosomes come from the same species.

**autoradiography** A technique in which radioactive molecules make their locations known by exposing photographic plates.

**autosomal set** A combination of nonsex chromosomes consisting of one from each homologous pair in a diploid species.

**autosomes** The nonsex chromosomes.

**autotrophs** Organisms that can utilize carbon dioxide as a carbon source.

**autozygosity** Homozygosity in which the two alleles are identical by descent (i.e., they are copies of an ancestral gene).

**auxotrophs** Organisms that have specific nutritional requirements.

**bacillus** A rod-shaped bacterium.

**backcross** The cross of an individual with one of its parents or with an organism with the same genotype as a parent.

**back mutation** The process that causes reversion. A change in a nucleotide pair in a mutant gene that restores the original sequence and hence the original phenotype.

**bacterial artificial chromosomes (BACs)** Artificial chromosomes used for sequencing that are derived from bacterial fertility factors (F plasmids).

**bacterial lawn** A continuous cover of bacteria on the surface of a growth medium.

**bacteriophages** Bacterial viruses.

**Balbiani rings** The larger polytene chromosomal puffs. Generally synonymous with *puffs*. *See* chromosome puffs.

**Barr body** Heterochromatic body (X chromosome) found in the nuclei of normal female mammals but absent in the nuclei of normal males.

**basal bodies** Microtubule organizing center for cilia and flagella, composed of centrioles.

**base excision repair** The DNA excision repair mechanism that replaces a nucleotide lacking its base with a complete nucleotide. Glycosylases or the environment create the AP (apurinic and apyrimidinic) nucleotides.

**base flipping** A process whereby enzymes gain access to bases within the DNA double helix by first flipping the bases out of the interior to the outside.

**basic/helix-loop-helix/leucine zipper** A motif of proteins that bind DNA that consists of a series of basic amino acids followed by a helix-loop-helix domain and then a leucine zipper. *See* helix-turn-helix motif; leucine zipper.

**Batesian mimicry** Form of mimicry in which an innocuous model gains protection by resembling a noxious or dangerous host.

**B DNA** The right-handed, double-helical form of DNA described by Watson and Crick.

**β-galactosidase** The enzyme that splits lactose into glucose and galactose (coded by a gene in the *lac* operon).

**β-galactoside acetyltransferase** An enzyme that is involved in lactose metabolism and encoded by a gene in the *lac* operon.

**β-galactoside permease** An enzyme involved in concentrating lactose in the cell (coded by a gene in the *1ac* operon).

**binary fission** Simple cell division in single-celled organisms.

**binomial expansion** The terms generated when a binomial expression is raised to a particular power.

**binomial theorem** The theorem that gives the terms of the expansion of a binomial expression raised to a particular power.

**biochemical genetics** The study of the relationships between genes and enzymes, specifically the role of genes in controlling the steps in biochemical pathways.

**bioinformatics** The science of storing, retrieving, and analyzing genomic data.

**biolistic** A method (*bio*logical bal*listic*) of transfecting cells by bombarding them with microprojectiles coated with DNA.

**biological species concept** The idea that organisms are classified in the same species if they are potentially capable of interbreeding and producing fertile offspring.

**bivalents** Structures, formed during prophase of meiosis I, consisting of the synapsed homologous chromosomes. Equivalent to a tetrad of chromatids.

**blastoderm** The outer layer of cells in an insect embryo after cleavage but before gastrulation. A syncitial blastoderm gives way to a cellular blastoderm when cell walls form.

**blunt-end ligation** The ligating or attaching of blunt-ended pieces of DNA by, for example, T4 DNA ligase. Used in creating hybrid vectors.

**bottleneck** A brief reduction in the size of a population, which usually leads to random genetic drift.

**bouquet stage** A stage during zygonema in which chromosome ends, attached to the nuclear membrane, come to lie near each other.

**branch migration** The process in which a crossover point between two duplexes slides along the duplexes.

**breakage and reunion** The general mode by which recombination occurs. DNA duplexes are broken and reunited in a crosswise fashion according to the Holliday model.

**breakage-fusion-bridge cycle** Damage that a dicentric chromosome goes through during each cell cycle.

**buoyant density of DNA** A measure of the density or size of DNA determined by the equilibrium point reached by DNA after density gradient centrifugation.

**calculus of the genes** Apparent calculation by the genes to determine when a particular altruistic behavior is beneficial to inclusive fitness and hence worth doing.

**cancer** An informal term for a diverse class of diseases marked by abnormal cell proliferation.

**cancer-family syndromes** Pedigree patterns in which unusually large numbers of blood relatives develop certain kinds of cancers.

**cap** A methylated guanosine added to the 5′ end of eukaryotic messenger RNA.

**capsid** The protein shell of a virus.

**capsomere** Protein clusters making up discrete subunits of a viral protein shell.

**carcinoma** Tumor arising from epithelial tissue (e.g., glands, breasts, skin, linings of the urogenital and respiratory systems).

**cassette mechanism** The mechanism by which homothallic yeast cells alternate mating types. The mechanism involves two silent transposons (cassettes) and a region where these cassettes can be expressed (cassette player).

**catabolite activator protein (CAP)** A protein that, when bound with cyclic AMP, can attach to sites on sugar-metabolizing operons to enhance transcription of these operons.

**catabolite repression** Repression of certain sugar-metabolizing operons in favor of glucose utilization when glucose is present in the environment of the cell.

**cDNA** *See* complementary DNA.

**cell cycle** The cycle of cell growth, replication of the genetic material, and nuclear and cytoplasmic division.

**cell-free system** A mixture of cytoplasmic components from cells, lacking nucleic acids and membranes. Used for in vitro protein synthesis and other purposes.

**cellular immunity** Immunity controlled by killer and helper T cells, which recognize infected body cells and either cause the infected cells to destroy internal invaders or destroy the infected cells directly.

**centimorgan** A chromosome-mapping unit. One centimorgan equals 1% recombinant offspring.

**central dogma** The original postulate of the way that information can be transferred between DNA, RNA, and protein, barring any transfer originating in protein.

**centric fragment** A piece of chromosome containing a centromere.

**centrioles** Cylindrical organelles, found in eukaryotes (except in higher plants), that reside in the centrosome. Also called basal bodies when they organize flagella or cilia.

**centromere** Constrictions in eukaryotic chromosomes on which the kinetochore lies. Also, the DNA sequence within the constriction that is responsible for appropriate function.

**centromeric fission** Creation of two chromosomes from one by splitting at the centromere.

**centrosome** The spindle-microtubule organizing center in eukaryotes except for those, such as fungi, that use spindle pole bodies to organize the spindles.

**chaperone** *See* molecular chaperone.

**Chargaff's rule** Chargaff's observation that in the base composition of DNA, the quantity of adenine equals the quantity of thymine, and the quantity of guanine equals the quantity of cytosine (equal purine and pyrimidine content).

**Charon phages** Phage lambda derivatives used as vehicles in DNA cloning.

**checkpoint** Used to describe points in the cell cycle that can be stopped if certain conditions are not met.

**chemiluminescent techniques** Techniques in which various molecules are made visible when exposed to ultraviolet or laser light if the molecules have a fluorescent tag.

**chiasmata** X-shaped configurations seen in tetrads during the later stages of prophase I of meiosis. They represent physical crossovers (singular: *chiasma*).

**chimeras** Individuals made up of two or more cell lines in which the cells originated in different zygotes. *See* mosaics.

**chimeric plasmid** Hybrid, or genetically mixed, plasmid used in DNA cloning.

**chi site** Sequence of DNA at which the RecBCD protein cleaves one of the DNA strands during recombination.

**chi-square distribution** The sampling distribution of the chi-square statistic. A family of curves whose shapes depend on degrees of freedom.

**chloroplast** The organelle that carries out photosynthesis and starch grain formation.

**chromatids** The subunits of a chromosome prior to anaphase of meiosis or mitosis. At anaphase of meiosis II or mitosis, when the sister chromatids separate, each chromatid becomes a chromosome.

**chromatin** The nucleoprotein material of the eukaryotic chromosome.

**chromatin assembly factors** Proteins involved in the construction of nucleosomes.

**chromatin remodeling** The change in the structure or positioning of nucleosomes, usually to allow transcription.

**chromatosome** The core nucleosome plus the Hl protein, a unit that includes approximately 168 base pairs of DNA.

**chromomeres** Dark regions of chromatin condensation in eukaryotic chromosomes at meiosis, mitosis, or endomitosis.

**chromosomal painting** A variant of the technique known as fluorescent in situ hybridization. Fluorescent dyes, attached to numerous nucleotide probes, give each human chromosome a different fluorescent signature.

**chromosomal theory of inheritance** The theory that chromosomes are linear sequences of genes.

**chromosome** The form of the genetic material in viruses and cells. A circle of DNA in most prokaryotes; a DNA or an RNA molecule in viruses; a linear nucleoprotein complex in eukaryotes.

**chromosome jumping** A technique for isolating clones from a genomic library that are not contiguous but skip a region between known points on the chromosome. This is usually done to bypass regions that are difficult or impossible to "walk" through or regions known not to be of interest.

**chromosome puffs** Diffuse, uncoiled regions in polytene chromosomes where transcription is actively taking place.

**chromosome walking** A technique for studying segments of DNA, larger than can be individually cloned, by using overlapping cloned DNA.

***cis*** Meaning "on the near side of"; refers to geometric configurations of atoms or mutants on the same chromosome.

***cis*-dominant** Mutants (e.g., of an operator) that control the functioning of genes on the same piece of DNA.

***cis-trans* complementation test** A mating test to determine whether two mutants on opposite chromosomes complement each other; a test for allelism.

**cistron** Term Benzer coined for the smallest genetic unit that exhibits the *cis-trans* position effect; synonymous with *gene*.

**cladogenesis** The evolutionary process whereby one species splits into two or more species. *See* anagenesis.

**classical linkage map** A chromosomal map, measured in centimorgans, based on genetic crosses to locate the relative distances between genes and their relative locations on chromosomes.

**clinal selection** Selection that changes gradually along a geographic gradient.

**clonal evolution theory** The theory that cancer develops from sequential changes (mutations) in the genome of a single cell.

**clone** A group of cells arising from a single ancestor.

**coccus** A spherical bacterium.

**coding strand** The DNA strand with the same sequence as the transcribed messenger RNA (given U in RNA and T in DNA). Compare with "anticoding strand."

**codominance** The relationship of alleles in a heterozygote that shows the individual expression of each allele in the phenotype.

**codon preference** The idea that for amino acids with several codons, one or a few are preferred and are used disproportionately. They would correspond with abundant transfer RNAs.

**codons** The sequences of three RNA or DNA nucleotides that specify either an amino acid or termination of translation.

**coefficient of coincidence** The number of observed double crossovers, divided by the number expected based on the independent occurrence of crossovers.

**coefficient of relationship, *r*** The proportion of alleles held in common by two related individuals.

**cohesin** Proteinaceous complex that holds sister chromatids together until anaphase of mitosis.

**cointegrate** A fusion of two elements. An intermediate structure in transposition.

**colicinogenic factors** *See* col plasmids.

**col plasmids** Plasmids that produce antibiotics (colicinogens) that the host uses to kill other strains of bacteria.

**combinatorial control** Transcriptional control in eukaryotes, which involves a large number of polypeptides, many of which recognize specific DNA sequences.

**common ancestry** The shared genetic inheritance of two individuals who are blood relatives. When two parents have a common ancestor, their offspring will be inbred.

**compensasome** The multisubunit dosage compensation complex in *Drosophila* with attendant RNAs.

**competence factor** A surface protein that binds extracellular DNA and enables the bacterial cell to be transformed.

**complementarity** The correspondence of DNA bases in the double helix so that adenine in one strand is opposite thymine in the other strand and cytosine in one strand is opposite guanine in the other. This relationship explains Chargaff's rule.

**complementary DNA (cDNA)** DNA synthesized by reverse transcriptase using RNA as a template.

**complementation** The production of the wild-type phenotype by a cell or an organism that contains two mutant genes. If complementation occurs, the mutants are almost certainly nonallelic.

**complementation group** Cistron (determined by the *cis-trans* complementation test).

**complete medium** A culture medium that is enriched to contain all of the growth requirements of a strain of organisms.

**component of fitness** A particular aspect in the life cycle of an organism upon which natural selection acts.

**composite transposon** A transposon constructed of two IS elements flanking a control region that frequently contains host genes.

**concordance** The amount of phenotypic similarity between individuals.

**condensin** A protein complex including SMC (structural maintenance of chromosomes) proteins needed for the condensation of interphase chromosomes to mitotic chromosomes.

**conditional-lethal mutant** A mutant that is lethal under one condition but not lethal under another.

**confidence limits** A statistical term for a pair of numbers that predicts the range of values within which a particular parameter lies.

**conjugation** A process whereby two cells come in contact and exchange genetic material. In prokaryotes, the transfer is a one-way process.

**consanguineous** Meaning "between blood relatives"; usually refers to inbreeding or incestuous matings.

**consensus sequence** A sequence of the common nucleotides found in many different DNA or RNA samples of homologous regions (e.g., promoters).

**conservative replication** A postulated mode of DNA replication in which an intact double helix acts as a template for a new double helix; known to be incorrect.

**conserved sequence** A sequence found in many different DNA or RNA samples (e.g., promoters) that is invariant in the sample.

**constitutive heterochromatin** Heterochromatin that surrounds the centromere. *See* satellite DNA.

**constitutive mutant** A mutant whose transcription is no longer under regulatory control.

**contigs** Genomic libraries of overlapping, contiguous clones that cover complete regions of a chromosome.

**continuous replication** In DNA, uninterrupted replication in the 5′ to 3′ direction using a 3′ to 5′ template.

**continuous variation** Variation measured on a continuum rather than in discrete units or categories (e.g., height in human beings).

**corepressor** The metabolite that when bound to the repressor (of a repressible operon) forms a functional unit that can bind to its operator and block transcription.

**correlation coefficient** A statistic that gives a measure of how closely two variables are related.

**cosmid** A hybrid plasmid that contains *cos* sites at each end. *Cos* sites are recognized during head filling of lambda phages. Cosmids are useful for cloning segments of foreign DNA up to 50 kb.

**cotransduction** The simultaneous transduction of two or more genes.

**coupling** Allelic arrangement in which mutants are on the same chromosome and wild-type alleles are on the homologue.

**covariance** A statistical value measuring the simultaneous deviations of $x$ and $y$ variables from their means.

**CpG islands** Stretches of CG repeats (in which CpG indicates sequential bases on the same strand of DNA, rather than a C-G base pair). These repeats, found in imprinting centers, are important in regulation.

**crisscross pattern of inheritance** The phenotypic pattern of inheritance controlled by X-linked recessive alleles.

**critical chi-square** A chi-square value for a given degree of freedom and probability level, to which we can compare an experimental chi-square.

**crossbreed** To facilitate fertilization between separate individuals.

**cross-fertilization** *See* crossbreed.

**crossing over** A process in which homologous chromosomes exchange parts by a breakage-and-reunion process.

**crossovers** *See* chiasmata.

**crossover suppression** The apparent lack of crossing over within an inversion loop in heterozygotes. Usually due to mortality of zygotes carrying defective crossover chromosomes rather than to actual suppression.

**C-value paradox** Structural and junk DNA create large eukaryotic genomes and large differences in DNA content between eukaryotic species.

**cyclic AMP** A form of AMP (adenosine monophosphate) used frequently as a second messenger in eukaryotic hormone nets and in catabolite repression in prokaryotes.

**cyclin** Family of proteins involved in cell cycle control.

**cyclin-dependent kinase (CDK)** Family of kinases (phosphorylating enzymes) that, when combined with cyclin, are active in controlling checkpoints in the cell cycle.

**cyclosome** *See* anaphase-promoting complex (APC).

**cytogenetics** The study of cells from the perspective of genetics. In practice, the study of changes in the gross structure and number of chromosomes in cells.

**cytokinesis** The division of the cytoplasm of a cell into two daughter cells. *See* karyokinesis.

**cytoplasmic inheritance** Extrachromosomal inheritance controlled by nonnuclear genomes.

**cytosine** *See* pyrimidines.

**cytotoxic T lymphocytes** T cells responsible for attacking host cells that have been infected with an invading bacterium or virus.

**dauermodification** The persistence for several generations of an environmentally induced trait.

**degenerate code** A code in which several code words have the same meaning. The genetic code is degenerate because many different codons may specify the same amino acid.

**degrees of freedom** An estimate of the number of independent categories in a particular statistical test or experiment.

**deletion chromosome** A chromosome with part deleted.

**deme** A locally interbreeding population.

**denatured** Loss of natural configuration (of a molecule) through heat or other treatment. Denatured DNA is single-stranded.

**denominator elements** Genes on the autosomes of *Drosophila* that regulate the sex switch (*sxl*) to the off condition (maleness). Refers to the denominator of the X/A genic balance equation.

**density-gradient centrifugation** A method of separating molecular entities by their differential sedimentation in a centrifugal gradient.

**depauperate fauna** A fauna, especially common on islands, lacking many species found in similar habitats.

**derepressed** The condition of an operon that is transcribing because repressor control has been lifted.

**deterministic** Referring to events that have no random or probabilistic aspects but proceed in a fixed, predictable fashion.

**development** The process of orderly change an individual goes through in the formation of structure.

**diakinesis** The final stage of prophase I of meiosis.

**dicentric chromosome** A chromosome with two centromeres.

**dictyotene** A prolonged diplonema of primary oocytes that can last many years.

**dideoxy method** A method of DNA sequencing that uses chain-terminating (dideoxy) nucleotides.

**dihybrid** An organism heterozygous at two loci.

**dimerization** The chemical union of two similar molecules.

**diploid** Having each chromosome in two copies per nucleus or cell.

**diplonema (diplotene stage)** The stage of prophase of meiosis I in which chromatids appear to repel each other.

**directed mutation** A form of mutation that appears to respond to the needs of the cell but may, in fact, be due to the cell's hypermutable state under duress.

**directional selection** A type of selection that removes individuals from one end of a phenotypic distribution and thus causes a shift in the distribution.

**disassortative mating** The mating of two individuals with dissimilar phenotypes.

**discontinuous replication** In DNA, the replication in short $5'$ to $3'$ segments, using the $5'$ to $3'$ strand as a template while going backward, away from the replication fork.

**discontinuous variation** Variation that falls into discrete categories (e.g., the green and yellow color of garden peas).

**discrete generations** Generations that have no overlapping reproduction. All reproduction takes place between individuals of the same cohort.

**dispersive replication** A postulated mode of DNA replication combining aspects of conservative and semiconservative replication; known to be incorrect.

**disruptive selection** A type of selection that removes individuals from the center of a phenotypic distribution and thus causes the distribution to become bimodal.

**D-loop** Configuration found during DNA replication of chloroplast and mitochondrial chromosomes wherein the origin of replication is different on the two strands. The first structure formed is a displacement loop, or D-loop.

**DNA cloning** *See* gene cloning.

**DNA-DNA hybridization** The process of taking DNA from the same or different sources and heating and then cooling it, causing double helices to re-form at homologous regions. This technique is useful for determining sequence similarities and degrees of repetitiveness among DNAs.

**DNA fingerprint** A pattern of bands created on an electrophoretic gel of a DNA digest probed for a variable locus.

**DNA glycosylases** Endonucleases that initiate excision repair at the sites of various damaged or improper bases in DNA.

**DNA gyrase** A topoisomerase that relieves supercoiling in DNA by creating a transient break in the double helix.

**DNA ligase** An enzyme that closes nicks or discontinuities in one strand of double-stranded DNA by creating an ester bond between adjacent $3'$-OH and $5'$-$PO_4$ ends on the same strand.

**DNA polymerase** One of several classes of enzymes that polymerize DNA nucleotides using single-stranded DNA as a template.

**DNA-RNA hybridization** The process of heating and then cooling a mixture of DNA and RNA so that the RNA can hybridize (form a double helix) with DNA with a complementary nucleotide sequence.

**docking protein** Responsible for attaching (docking) a ribosome to a membrane by interacting with a signal particle attached to a ribosome destined to be membrane bound.

**dominant** An allele that expresses itself even when heterozygous. Also, the trait controlled by that allele.

**dosage compensation** A mechanism by which species with sex chromosomes ensure that one sex does not have differential activity of alleles on the sex chromosomes.

**dot blotting** A blotting technique, used on DNA already cloned, that eliminates the electrophoretic separation step. Autoradiographs reveal dots rather than bands on a gel, indicating a probed sequence.

**double digest** The product formed when two different restriction endonucleases act on the same segment of DNA.

**double helix** The normal structural configuration of DNA consisting of two helices rotating about the same axis.

**downstream** A convention on DNA related to the position and direction of transcription by RNA polymerase ($5'\rightarrow3'$). Downstream (in the $3'$ direction) is in the direction of transcription, whereas upstream (in the $5'$ direction) is in the direction from which the polymerase has come.

**downstream promoter element (DPE)** A consensus sequence at about $+28$ to $+34$ of RNA polymerase II promoters that have initiator elements but not TATA boxes.

**dyad** Two sister chromatids attached to the same centromere.

**dynein** Microtubule motor protein.

**dysplasia** Excessive cell growth that involves pathological changes to the cells and their nuclei.

**electrophoresis** The separation of molecular entities by electric current.

**electroporation** A technique for transfecting cells by applying a high-voltage electric pulse.

**elongation complex** The form of RNA polymerase II that actively carries out basal transcription.

**elongation factors (EF-Ts, EF-Tu, EF-G)** Proteins necessary for the proper elongation and translocation processes during translation at the ribosome in prokaryotes. Replaced by eEFIα and eEFIβγ in eukaryotes.

**endogenote** Bacterial host chromosome.

**endomitosis** Chromosomal replication without nuclear or cellular division that results in cells with many copies of the same chromosome, such as in the salivary glands of *Drosophila*.

**endonucleases** Enzymes that make nicks internally in the backbone of a polynucleotide. They hydrolyze internal phosphodiester bonds.

**enhancer** A eukaryotic DNA sequence that increases transcription of a gene by binding specific transcription factors.

**enriched medium** *See* complete medium.

**enzyme** Protein catalyst.

**epigenetic effect** An environmentally induced change in the genetic material that does not cause a change in base pairs. Generally, a phenomenon of differential expression of alleles of a locus depending on the parent of origin. Also applied to an effect in proteins.

**epistasis** The masking of the action of alleles of one gene by allelic combinations of another gene.

**equational division** A division, such as the second meiotic division, that does not reduce chromosomal numbers.

**E (exit) site** Site on the ribosome that depleted transfer RNAs pass through during ejection.

**euchromatin** Regions of eukaryotic chromosomes that are diffuse during interphase. Presumably the actively transcribing DNA of the chromosomes.

**eugenics** A social movement designed to improve humanity by encouraging those with beneficial traits to breed and discouraging those with undesirable traits from breeding.

**eukaryotes** Organisms with true nuclei.

**euploidy** The condition of a cell or organism that has one or more complete sets of chromosomes.

**evolution** A change in phenotypic frequencies in a population.

**evolutionary rates** The rate of divergence between taxonomic groups, measurable as number of amino acid substitutions per million years.

**excision repair** A process whereby cells remove part of a damaged DNA strand and replace it through DNA synthesis, using the undamaged strand as a template.

**exconjugant** Each of the two cells that separate after conjugation has taken place.

**exogenote** DNA that a bacterial cell has taken up through one of its sexual processes.

**exon** In a gene that has intervening sequences (introns), a region that is actually exported from the nucleus to be expressed or become part of a transfer or ribosomal RNA.

**exon shuffling** The hypothesis put forward by Walter Gilbert that exons code for the functional units of a protein, and that the evolution of new genes proceeds by recombination or the exclusion of exons.

**exonucleases** Enzymes that digest nucleotides from the ends of polynucleotide molecules. They hydrolyze phosphodiester bonds of terminal nucleotides.

**experimental design** A branch of statistics that attempts to outline the way in which experiments should be carried out so the data gathered has statistical value.

**expression vector** A hybrid vector (plasmid) that expresses its cloned genes.

**expressivity** The degree of expression of a genetically controlled trait.

**$F_1$** *See* filial generation.

**factorial** The product of all integers from the specified number down to one (unity).

**Fanconi's anemia** A disease in human beings with a syndrome of congenital malformations; associated with various cancers.

**fate map** A map of the developmental fate of a zygote or early embryo showing the adult organs that will develop from a given position on the zygote or early embryo.

**F-duction** *See* sexduction.

**fecundity selection** The forces causing one genotype to be more fertile than another genotype.

**feedback inhibition** A posttranslational control mechanism in which the end product of a biochemical pathway inhibits the activity of the first enzyme of the same pathway.

**fertility factor** The plasmid that allows a prokaryote to engage in conjugation with, and pass DNA into, an $F^-$ cell.

**F factor** *See* fertility factor.

**filial generation** Offspring generation. $F_1$ is the first offspring, or filial, generation; $F_2$ is the second; and so on.

**fimbriae** *See* pili.

**first-division segregation (FDS)** The allelic arrangement of ordered spores that indicates the lack of recombination between a locus and its centromere.

**fitness, *W*** The relative reproductive success of a genotype as measured by survival, fecundity, or other life-history parameters.

**5′ untranslated region (5′ UTR)** *See* leader.

**floral meristem** The shoot apical meristem sets aside this tissue that gives rise to flowers.

**floral-meristem identity genes** At least five genes known to establish the identity of the floral meristem.

**fluctuation test** An experiment by Luria and Delbrück that compared the variance in number of mutations among small cultures with that among subsamples of a large culture to determine the mechanism of inherited change in bacteria.

**fluorescent in situ hybridization (FISH)** A technique in which a fluorescent dye is attached to a nucleotide probe that then binds to a specific site on a chromosome and makes itself visible by its fluorescence.

**Fokker-Planck equation** An equation that describes diffusion processes. It is used by population geneticists to describe random genetic drift.

**footprinting** A technique to determine the length of nucleic acid in contact with a protein. While in contact, the free DNA is digested. The remaining DNA is then isolated and characterized.

**founder effect** Genetic drift observed in a population founded by a small, nonrepresentative sample of a larger population.

**F-pili** Sex pili. Hairlike projections on an $F^+$ or Hfr bacterium involved in anchorage during conjugation.

**fragile site** A chromosomal region that has a tendency to break.

**fragile-X syndrome** The most common form of inherited mental retardation. Named for its association with an X chromosome with a tip that breaks or appears uncondensed. Inheritance involves imprinting.

**frameshift** A mutation in which there is an addition or deletion of nucleotides that causes the codon reading frame to shift.

**free energy of activation ($\Delta G^{\ddagger}$)** Energy needed to initiate a chemical reaction.

**frequency-dependent selection** A selection whereby a genotype is at an advantage when rare and at a disadvantage when common.

**functional alleles** Mutations that fail to complement each other in a *cis-trans* complementation test.

**fundamental number (NF)** The number of chromosome arms in a somatic cell of a particular species.

**gamete** A germ cell having a haploid chromosomal complement. Gametes from parents of opposite sexes fuse to form zygotes.

**gametic selection** The forces acting to cause differential reproductive success of one allele over another in a heterozygote.

**gametophyte** The haploid stage of a plant life cycle that produces gametes (by mitosis). It alternates with a diploid, sporophyte generation.

**G-bands** Eukaryotic chromosomal bands produced by treatment with Giemsa stain.

**gene** Inherited determinant of the phenotype. *See* cistron; locus.

**gene amplification** A process or processes by which the cell increases the number of repeats of a particular gene within the genome.

**gene cloning** Production of large numbers of a piece of DNA after that piece of DNA is inserted into a vector and taken up by a cell. Cloning occurs as the vector replicates.

**gene conversion** In Ascomycete fungi, a 2:2 ratio of alleles is expected after meiosis, yet a 3:1 ratio is sometimes observed. The gene conversion mechanism is explained by repair of heteroduplex DNA produced by recombination.

**gene family** A group of genes that has arisen by duplication of an ancestral gene. The genes in the family may or may not have diverged.

**gene flow** The movement of genes from one population to another by interbreeding between individuals in the two populations.

**gene pool** All of the alleles available among the reproductive members of a population from which gametes can be drawn.

**generalized transduction** Form of transduction in which any region of the host genome can be transduced. *See* specialized transduction.

**general transcription factors** Eukaryotic proteins that form part of the RNA polymerase holoenzymes.

**genetic code** The linear sequences of nucleotides that specify the amino acids during the process of translation at the ribosome.

**genetic engineering** Popular term for recombinant DNA technology. *See* recombinant DNA technology.

**genetic load** The relative decrease in the mean fitness of a population due to the presence of genotypes that have less than the highest fitness.

**genetic polymorphism** The occurrence together in the same population of more than one allele at the same locus, with the least frequent allele occurring more frequently than can be accounted for by mutation.

**genic balance theory** Bridges's theory that the sex of a fruit fly is determined by the relative number of X chromosomes and autosomal sets.

**genome** The entire genetic complement of a prokaryote or virus or the haploid genetic complement of a eukaryote.

**genomic equivalence** The concept that differentiated cells in a eukaryotic organism have identical genetic contents.

**genomic library** A set of cloned fragments making up the entire genome of an organism or species.

**genomics** The study of the mapping and sequencing of genomes. Bioinformatics is the science of mining the data from these DNA sequences obtained from sequencing.

**genophore** The chromosome (genetic material) of prokaryotes and viruses.

**genotype** The genes that an organism possesses.

**Giemsa stain** A complex of stains specific for the phosphate groups of DNA.

**Goldstein-Hogness box** *See* TATA box.

**green fluorescent protein** A reporter system that uses the gene from a jellyfish that specifies a protein that fluoresces green when ultraviolet light is shined on it, indicating the success of a transfection experiment.

**group I introns** Self-splicing introns that require a guanine-containing nucleotide for splicing; the intron is released in a linear form.

**group II introns** Self-splicing introns that do not require an external nucleotide for splicing; the intron is released in a lariat form.

**group selection** Selection for traits that would be beneficial to a population at the expense of the individual possessing the trait.

**G-tetraplex** A structure of four guanines that can base pair to form a planar structure that may be involved in novel structures at the end of eukaryotic chromosomes.

**guanine** *See* purines.

**guide RNA (gRNA)** RNA that guides the insertion of uridines (RNA editing) into messenger RNAs in trypanosomes. Found in transcripts from minicircles and maxicircles of DNA in kinetoplasts.

**gynandromorphs** Mosaic individuals having simultaneous aspects of both the male and the female phenotype.

**hammerhead ribozyme** A catalytic RNA, shaped like a hammerhead, capable of splitting other RNA molecules with appropriate complementary sequences.

**haplodiploidy** The sex-determining mechanism found in some insect groups among which males are haploid and females are diploid.

**haploid** The state of having one copy of each chromosome per nucleus or cell.

**HAT medium** A selection medium for hybrid cell lines; contains hypoxanthine, aminopterin, and thymidine. HPRT$^+$ TK$^+$ cell lines can survive in this medium.

**heat shock proteins** Proteins that appear in a cell after the cell has been subjected to elevated temperatures.

**helicase** A protein that unwinds DNA, usually at replicating Y-junctions.

**helix-turn-helix motif** Configuration found in DNA-binding proteins consisting of a recognition helix and a stabilizing helix. separated by a short turn.

**hemizygous** The condition of loci present in only one copy in a diploid organism, such as loci on the X chromosome of the heterogametic sex of a diploid species.

**heritability** A measure of the degree to which the variance in the distribution of a phenotype is due to genetic causes. In the broad sense, it is measured by the total genetic variance divided by the total phenotypic variance. In the narrow sense, it is measured by the genetic variance due to additive genes divided by the total phenotypic variance.

**hermaphrodite** An individual with both male and female genitalia.

**heterochromatin** Chromatin that remains tightly coiled (and darkly staining) throughout the cell cycle.

**heteroduplex DNA** *See* hybrid DNA.

**heterogametic** The sex with heteromorphic sex chromosomes; during meiosis, it produces different kinds of gametes in accordance with these sex chromosomes.

**heterogeneous nuclear mRNA (hnRNA)** The original RNA transcripts found in eukaryotic nuclei before posttranscriptional modifications.

**heterokaryon** A cell that contains two or more nuclei from different origins.

**heteromorphic chromosome pair** Members of a homologous pair of chromosomes that are not morphologically identical (e.g., the sex chromosomes).

**heteroplasmy** The existence within an organism of genetic heterogeneity within the populations of mitochondria or chloroplasts.

**heterothallic** A botanical term used for organisms in which the two sexes reside in different individuals.

**heterotrophs** Organisms that require an organic form of carbon as a carbon source.

**heterozygote** A diploid or polyploid with different alleles at a particular locus.

**heterozygote advantage** A selection model in which heterozygotes have the highest fitness.

**heterozygous DNA** *See* hybrid DNA.

**Hfr** High frequency of recombination. A strain of bacteria that has incorporated an F factor into its chromosome and can then transfer the chromosome during conjugation.

**histone acetyl transferases (HATs)** Proteins that remodel chromatin by acetylating histones.

**histones** Arginine- and lysine-rich basic proteins making up a substantial portion of eukaryotic nucleoprotein.

**hnRNA** *See* heterogeneous nuclear mRNA.

**Hogness box** *See* TATA box.

**holandric trait** Trait controlled by a locus found only on the Y chromosome. Involves father-to-son transmission.

**Holliday junction** A junction point between two cross-linked DNA double helices. It is an intermediate stage in DNA recombination.

**holoenzyme** The complete enzyme, including all subunits. Often used in reference to RNA and DNA polymerases.

**homeo box** A consensus sequence of about 180 base pairs discovered in homeotic genes in *Drosophila*. Also found in other developmentally important genes from yeast to human beings.

**homeo domain** The sixty amino acid polypeptide translated from the homeo box.

**homeotic gene** Gene that controls the developmental fate of a cell type; mutations of the homeotic gene cause one cell type to follow the developmental pathway of another cell type.

**homogametic** The sex with homomorphic sex chromosomes; it produces only one kind of gamete in regard to the sex chromosomes.

**homologous chromosomes** Members of a pair of essentially identical chromosomes that synapse during meiosis.

**homologous recombination** Breakage and reunion between homologous lengths of DNA mediated by RecA and RecBCD.

**homomorphic chromosome pairs** Morphologically identical members of a homologous pair of chromosomes.

**homoplasmy** The existence within an organism of only one type of plastid; usually referring to the genetic identity of mitochondria or chloroplasts.

**homothallic** A botanical term used for groups whose individuals are not of different sexes.

**homozygote** A diploid or a polyploid with identical alleles at a locus.

**humoral immunity** Immunity due to antibodies in the serum and lymph.

**H-Y antigen** The histocompatibility Y-antigen, a protein found on the cell surfaces of male mammals.

**hybrid** Offspring of unlike parents.

**hybrid DNA** DNA whose two strands have different origins.

**hybridoma** A cell resulting from the fusion of a spleen cell and a multiple myeloma cell. These cells can be maintained indefinitely in cell culture, in which they produce monoclonal antibodies.

**hybrid plasmid** A plasmid that contains an inserted piece of foreign DNA.

**hybrid vector** *See* hybrid vehicle.

**hybrid vehicle** A plasmid or phage containing an inserted piece of foreign DNA.

**hybrid zone** Geographical region in which previously isolated populations that have evolved differences come into contact and form hybrids.

**hyperplasia** Excessive cell growth that does not involve pathological changes to the cells.

**hypervariable loci** Loci with many alleles; especially those whose variation is due to variable numbers of tandem repeats.

**hypostatic gene** A gene whose expression is masked by an epistatic gene.

**identity by descent** The state of two alleles when they are identical copies of the same ancestral allele (autozygous).

**idiogram** A photograph or diagram of the chromosomes of a cell arranged in an orderly fashion. *See* karyotype.

**idiotypic variation** Variation in the variable parts of immunoglobulin genes.

**idling reaction** The production of guanosine tetraphosphate (3′-ppGpp-5′) by the stringent factor when a ribosome encounters an uncharged transfer RNA in the A site.

**immunity** The ability of an organism to resist infection.

**immunoglobulins (Igs)** Specific proteins produced by derivatives of B lymphocytes that protect an organism from antigens.

**imprinting** *See* molecular imprinting.

**imprinting center (IC)** A region responsible for the control of imprinting. The imprinting mark is almost certainly DNA methylation, which is able to turn off gene transcription.

**inbreeding** The mating of genetically related individuals.

**inbreeding coefficient, F** The probability of autozygosity.

**inbreeding depression** A depression of vigor or yield due to inbreeding.

**incestuous** A mating between blood relatives who are more closely related than the law of the land allows.

**inclusive fitness** The expansion of the concept of the fitness of a genotype to include benefits accrued to relatives of an individual since relatives share parts of their genomes. An apparently altruistic act toward a relative may thus enhance the fitness of the individual performing the act.

**incomplete dominance** The situation in which both alleles of the heterozygote influence the phenotype. The phenotype is usually intermediate between the two homozygous forms.

**independent assortment, rule of** Mendel's second rule, describing the independent segregation of alleles of different loci.

**inducible system** A system (a coordinated group of enzymes involved in a catabolic pathway) is inducible if the metabolite it works upon causes transcription of the genes controlling these enzymes. These systems are primarily prokaryotic operons.

**induction** In regard to temperate phages, the process of causing a prophage to become virulent.

**informatics** *See* bioinformatics.

**initiation codon** The messenger RNA sequence AUG, which specifies methionine, the first amino acid used in the translation process. (Occasionally, GUG is recognized as an initiation codon.)

**initiation complex** The complex formed for initiation of translation. It consists of the 30S ribosomal subunit, messenger RNA, N-formyl-methionine transfer RNA, and three initiation factors.

**initiation factors (IFI, IF2, IF3)** Proteins (prokaryotic with eukaryotic analogues) required for the proper initiation of translation.

**initiator element (inr)** A CT-rich area found in RNA polymerase II promoters without TATA boxes.

**initiator proteins** Proteins that recognize the origin of replication on a replicon and take part in primosome construction.

**insertion mutagenesis** Change in gene action due to an insertion event that either changes a gene directly or disrupts control mechanisms.

**insertion sequences (IS)** Small, simple transposons. *See* transposable genetic element.

**inside marker** The middle locus of three linked loci.

**intercalary heterochromatin** Heterochromatin, other than centromeric heterochromatin, dispersed throughout eukaryotic chromosomes.

**intergenic suppression** A mutation at a second locus that apparently restores the wild-type phenotype to a mutant at a first locus.

**interkinesis** The abbreviated interphase that occurs between meiosis I and II. No DNA replication occurs here.

**internal ribosome entry site** Sequence in eukaryotic messenger RNAs that allows ribosomes to initiate translation at a point other than the 5′ cap.

**interphase** The metabolically active, nondividing stage of the cell cycle.

**interpolar microtubules** Microtubules extending from one pole of the spindle and overlapping spindle fibers from the other pole, but not in contact with kinetochores.

**interrupted mating** A mapping technique in which bacterial conjugation is disrupted after specified time intervals.

**intersex** An organism with external sexual characteristics of both sexes.

**intervening sequences (introns)** DNA sequences within a gene that are transcribed but removed prior to translation.

**intra-allelic complementation** The restoration of activity in an enzyme made of subunits in a heterozygote of two mutants that, when homozygous, are not active. Caused by the interaction of the subunits in the protein.

**intragenic suppression** A second change within a mutant gene that results in an apparent restoration of the original phenotype.

**intron** *See* intervening sequences.

**inversion** The replacement of a section of a chromosome in the reverse orientation.

**inverted repeat sequence** A nucleotide sequence read in opposite orientations on the same double helix.

**in vitro** Biological or chemical work done in the test tube (literally, "in glass") rather than in living systems.

**IS elements** *See* insertion sequences.

**isochromosome** A chromosome with two genetically and morphologically identical arms.

**isozymes** Different electrophoretic forms of the same enzyme. Unlike allozymes, isozymes are due to differing subunit configurations rather than allelic differences.

**junctional diversity** Variability in immunoglobulins caused by variation in the exact crossover point during V-J, V-D, and D-J joining.

**kappa particles** The bacterialike particles that give a *Paramecium* the killer phenotype.

**karyokinesis** The process of nuclear division. *See* cytokinesis.

**karyotype** The chromosome complement of a cell. *See* idiogram.

**kinesin** Microtubule motor protein.

**kinetochore** The chromosomal attachment point for the spindle fibers, located on the centromere.

**kinetochore microtubules** Microtubules radiating from the centrosome and attached to kinetochores of chromosomes during mitosis and meiosis.

**kin selection** The mode of natural selection that acts on an individual's inclusive fitness.

**Klenow fragment** Proteolytic fragment—obtained by treatment with a protease, or protein-cleaving enzyme—of *E. coli* DNA polymerase I with both 5′→3′ polymerase activity and 3′→5′ exonuclease activity. (It has been studied extensively because it has been easier for X-ray crystallographers and biochemists to work with this fragment than with the whole enzyme.)

**knockout mice** Transgenic mice that have been made homozygous for a nonfunctioning allele at a particular locus.

***lac* operon** The inducible operon, including three loci involved in the uptake and breakdown of lactose.

**ladder gel** *See* stepladder gel.

**lagging strand** Strand of DNA being replicated discontinuously.

**lampbrush chromosomes** Chromosomes of amphibian oocytes having loops that are suggestive of a lampbrush.

**leader** The length of messenger RNA from the 5′ end to the initiation codon, AUG.

**leader peptide gene** A small gene within the attenuator control region of a repressible amino acid operon. Translation of the gene tests the concentration of amino acids in the cell.

**leader transcript** The messenger RNA transcribed by the attenuator region of a repressible amino acid operon. The transcript is capable of several alternative stem-loop structures, depending on the translation of a short leader peptide gene.

**leading strand** Strand of DNA being replicated continuously.

**leptonema (leptotene stage)** The first stage of prophase I of meiosis, in which chromosomes become distinct.

**lethal-equivalent alleles** Alleles whose summed effect is that of lethality—for example, four alleles, each of which would be lethal 25% of the time (or to 25% of their bearers), are equivalent to one lethal allele.

**leucine zipper** Configuration of a DNA-binding protein in which leucine residues on two helices interdigitate, in zipper fashion, to stabilize the protein.

**leukemia** Cancer of the bone marrow resulting in excess production of leukocytes.

**level of significance** The probability value in statistics used to reject the null hypothesis.

**linkage** The association of loci on the same chromosome.

**linkage disequilibrium** The condition among alleles at different loci such that allelic combinations in a gamete do not follow the product rule of probability.

**linkage equilibrium** The condition among alleles at different loci such that any allelic combination in a gamete occurs as the product of the frequencies of each allele at its own locus.

**linkage groups** Associations of loci on the same chromosome. In a species, there are as many linkage groups as there are homologous pairs of chromosomes.

**linkage number** The number of times one strand of a helix coils about the other.

**linker** A small segment of DNA that contains a restriction site. It can be added to blunt-ended DNA to give that DNA a particular restriction site for cloning.

**liposomes** Transfecting DNA is delivered to target cells by way of these membrane-bound vesicles.

**locus** The position of a gene on a chromosome (plural: *loci*).

***lod* score method** A technique (*l*ogarithmic *od*ds) for determining the most likely recombination frequency between two loci from pedigree data.

**long interspersed elements (LINEs)** Sequences of DNA, up to seven thousand base pairs in length, interspersed in eukaryotic chromosomes in many copies.

**lymphoma** Cancer of the lymph nodes and spleen that causes excessive production of lymphocytes.

**Lyon hypothesis** The hypothesis that suggests that the Barr body is an inactivated X chromosome.

**lysate** The contents released from a lysed cell.

**lysis** The breaking open of a cell by the destruction of its wall or membrane.

**lysogenic** The state of a bacterial cell that has an integrated phage (prophage) in its chromosome.

**major histocompatibility complex** A group of highly polymorphic genes whose products appear on the surfaces of cells, imparting to them the property of "self" (belonging to that organism). Some other functions are also involved.

**mapping** The process of locating the positions of genes on chromosomes.

**mapping function** The mathematical relationship between measured map distance in a given experiment and the actual recombination frequency.

**map unit** The distance equal to 1% recombination between two loci.

**mate-killer** A *Paramecium* phenotype induced by intracellular bacterialike mu particles.

**maternal effect** The effect of the maternal parent's genotype on the phenotype of her offspring.

**maternal-effect gene** A gene expressed in maternal tissue that influences a developing embryo.

**mating type** In many species of microorganisms, individuals can be divided into two mating types. Mating can take place only between individuals of opposite mating types due to the interaction of cell surface components.

**maturation-promoting factor (MPF)** A protein complex of cyclin B and p34*cdc2* that initiates mitosis during the cell cycle. Also called the mitosis-promoting factor.

**mean** The arithmetic average; the sum of the data divided by the sample size.

**mean fitness of the population, $\overline{W}$** The sum of the fitnesses of the genotypes of a population weighted by their proportions; hence, a weighted mean fitness.

**meiosis** The nuclear process in diploid eukaryotes that results in gametes or spores with only one member of each original homologous pair of chromosomes per nucleus.

**meiotic drive** *See* gametic selection.

**merozygote** A bacterial cell having a second copy of a particular chromosomal region in the form of an exogenote.

**messenger RNA (mRNA)** A complementary copy of a gene that is translated into a polypeptide at the ribosome.

**metacentric chromosome** A chromosome whose centromere is located in the middle.

**metafemale** A fruit fly with an X/A ratio greater than unity.

**metagon** An RNA necessary for the maintenance of mu particles in *Paramecium*.

**metamale** A fruit fly with X/A ratio below 0.5.

**metaphase** The stage of mitosis or meiosis in which spindle fibers are attached to kinetochores, and the chromosomes are positioned in the equatorial plane of the cell.

**metaphase plate** The plane of the equator of the spindle into which chromosomes are positioned during metaphase.

**metastasis** The migration of cancerous cells to other parts of the body.

**metrical variation** *See* continuous variation.

**microsatellite DNA** Repeats of very short sequences of DNA, such as CACACACA, dispersed throughout the eukaryotic genome. The loci can be studied by polymerase chain reaction amplification.

**microtubule organizing center** Active center from which microtubules are organized. The spindle is organized by the centrosome, which may or may not contain a centriole.

**microtubules** Hollow cylinders made of the protein tubulin (α and β subunits) that make up, among other things, the spindle fibers.

**mimicry** A phenomenon in which an individual gains an advantage by looking like the individuals of a different species.

**minimal medium** A culture medium for microorganisms that contains the minimal necessities for growth of the wild-type.

**mismatch repair** A form of excision repair initiated at the sites of mismatched bases in DNA.

**missense mutation** Mutations that change a codon for one amino acid into a codon for a different amino acid.

**mitochondrion** The eukaryotic cellular organelle in which the Krebs cycle and electron transport reactions take place.

**mitosis** The nuclear division producing two daughter nuclei identical to the original nucleus.

**mitosis-promoting factor** *See* maturation-promoting factor (MPF).

**mitotic apparatus** *See* spindle.

**mixed families** Groups of four codons sharing their first two bases and coding for more than one amino acid.

**modern linkage map** A chromosomal map based on the positions of RFLP markers along its length.

**molecular chaperone** A protein that aids in the folding of a second protein. The chaperone prevents proteins from forming structures that would be inactive.

**molecular evolutionary clock** A measurement of evolutionary time in nucleotide substitutions per year.

**molecular imprinting** The phenomenon in which there is differential expression of a gene depending on whether it was maternally or paternally inherited.

**molecular mimicry** The situation in which one type of molecule resembles another type in order to function. For example, the prokaryotic ribosomal release factors, RF1 and RF2, mimic the structure of a transfer RNA.

**monocistronic** Usually referring to a messenger RNA that carries the information for only one gene (cistron).

**monoclonal antibody** The antibody from a clone of cells producing the same antibody. An individual with multiple myeloma usually produces monoclonal antibodies.

**monohybrids** Offspring of parents that differ in only one genetic characteristic. Usually implies heterozygosity at a single locus under study.

**monosomic** A diploid cell missing a single chromosome.

**monovalent** A single chromosome composed of two sister chromatids. Equivalent to a dyad.

**morphogen** A substance transported into or produced in a developing embryo that diffuses to form a gradient that helps determine cell differentiation.

**morphological species concept** The idea that organisms are classified in the same species if they appear similar.

**mosaicism** The condition of being a mosaic. *See* mosaics.

**mosaics** Individuals made up of two or more cell lines in which the cells originated in the same zygote.

**mRNA** *See* messenger RNA (mRNA).

**Müllerian mimicry** A form of mimicry in which noxious species evolve to resemble each other.

**multihybrid** An organism heterozygous at numerous loci.

**multinomial expansion** The terms generated when a multinomial is raised to a power.

**mu particles** Bacterialike particles found in the cytoplasm of *Paramecium* that impart the mate-killer phenotype.

**mutability** The ability to change.

**mutants** Alternative phenotypes to the wild-type; the phenotypes produced by alternative alleles.

**mutation** The process by which a gene or chromosome changes structurally; the end result of this process.

**mutation rate** The proportion of mutations per cell division in bacteria or single-celled organisms or the proportion of mutations per gamete in higher organisms.

**mutator mutations** Mutations of DNA polymerase that increase the overall mutation rate of a cell or of an organism.

**muton** A term Benzer coined for the smallest mutable site within a cistron.

**natural selection** The process in nature whereby one genotype leaves more offspring than another genotype because of superior life history attributes such as survival or fecundity.

**negative interference** The phenomenon whereby a crossover in a particular region facilitates the occurrence of other apparent crossovers in the same region of the chromosome.

**N-end rule** The life span of a protein is determined by its amino-terminal (N-terminal) amino acid.

**neo-Darwinism** The merger of classical Darwinian evolution with population genetics.

**neoplasm** New growth of abnormal tissue.

**neutral gene hypothesis** The hypothesis that most genetic variation in natural populations is not maintained by natural selection.

**NF** *See* fundamental number.

**nickase** *See* DNA gyrase.

**noncoding strand** *See* anticoding strand.

**nondisjunction** The failure of a pair of homologous chromosomes to separate properly during meiosis.

**nonhistone proteins** The proteins remaining in chromatin after the histones are removed.

**nonparental ditype (NPD)** A spore arrangement in Ascomycete fungi that contains only the two recombinant-type ascospores (assuming two segregating loci).

**nonparentals** *See* recombinants.

**nonrecombinants** In mapping studies, offspring that have alleles arranged as in the original parents.

**nonsense codon** One of the messenger RNA sequences (UAA, UAG, UGA) that signals the termination of translation.

**nonsense mutations** Mutations that change a codon for an amino acid into a nonsense codon.

**normal distribution** Any of a family of bell-shaped frequency curves whose relative positions and shapes are defined on the basis of the mean and standard deviation.

**northern blotting** A gel transfer technique used for RNA. *See* Southern blotting.

**N segments** Sequences of nucleotides added in a template-free fashion at the joining junctions of heavy-chain antibody genes.

**nuclease-hypersensitive site** A region of a eukaryotic chromosome that is specifically vulnerable to nuclease attack because it is not wrapped as nucleosomes.

**nucleolar organizer** The chromosomal region around which the nucleolus forms; site of tandem repeats of the major ribosomal RNA gene.

**nucleolus** The globular, nuclear organelle formed at the nucleolar organizer. Site of ribosome construction.

**nucleoprotein** The substance of eukaryotic chromosomes consisting of proteins and nucleic acids.

**nucleoside** A sugar-base compound that is a nucleotide precursor. Nucleotides are nucleoside phosphates.

**nucleosomes** Arrangements of DNA and histones forming regular spherical structures in eukaryotic chromatin.

**nucleotide** Subunits that polymerize into nucleic acids (DNA or RNA). Each nucleotide consists of a nitrogenous base, a sugar, and one or more phosphate groups.

**nucleotide excision repair** The DNA excision repair mechanism responsible for repairing thymine dimers and other lesions. Enzymes excise a short segment of one of the DNA strands and then repair and ligate the DNA.

**null hypothesis** The statistical hypothesis that there are no differences between observed data and those data expected based on the assumption of no experimental effect.

**nullisomic** A diploid cell missing both copies of the same chromosome.

**numerator elements** Genes on the X chromosome in *Drosophila* that regulate the sex switch gene (*sxl*) to the on condition (femaleness). Refers to the numerator of the X/A genic balance equation.

**nutritional-requirement mutants** *See* auxotrophs.

**Okazaki fragments** Segments of newly replicated DNA produced during discontinuous DNA replication.

**oncogene** Genes capable of transforming a cell. They are found in the active state in retroviruses and transformed cells and in the inactive state in nontransformed cells, in which they are called proto-oncogenes.

**one-gene-one-enzyme hypothesis** Hypothesis of Beadle and Tatum that states that one gene controls the production of one enzyme. Later modified to the concept that one cistron controls the production of one polypeptide.

**oogenesis** The process of ovum formation in female animals.

**oogonia** Cells in females that produce primary oocytes by mitosis.

**open reading frames (ORFs)** Sequence of codons between the initiation and termination codons in a gene.

**operator** A DNA sequence recognized by a repressor protein or repressor-corepressor complex. When the operator is complexed with the repressor, transcription is prevented.

**operon** A sequence of adjacent genes all under the transcriptional control of the same operator.

**origin recognition complex (ORC)** A complex of six proteins that bind to the eukaryotic autonomously replicating sequences (ARS). Needed for the initiation of DNA replication in concert with other proteins.

**outbreeding** The mating of genetically unrelated individuals.

**ovum** Egg. The one functional product of each meiosis in female animals.

**pachynema (pachytene stage)** The stage of prophase I of meiosis in which chromatids are first distinctly visible.

**palindrome** A sequence of words, phrases, or nucleotides that reads the same regardless of the direction from which one starts; the sites of recognition of type II restriction endonucleases.

**panmictic** Referring to unstructured (random mating) populations.

**paracentric inversion** A chromosomal inversion that does not include the centromere.

**paramecin** A toxin liberated by a "killer" *Paramecium*.

**parameters** Measurements of attributes of a population; denoted by Greek letters.

**parapatric speciation** Speciation in which reproductive isolating mechanisms evolve when a population enters a new niche or habitat within the range of the parent species.

**parasegment** The first series of segments that form in a developing insect embryo; they form after about 5.5 hours in the developing *Drosophila* embryo.

**parental ditype (PD)** A spore arrangement in Ascomycete fungi that contains only the two nonrecombinant-type ascospores.

**parental imprinting** *See* molecular imprinting.

**parentals** *See* nonrecombinants.

**parthenogenesis** The development of an individual from an unfertilized egg that did not arise by meiotic chromosomal reduction.

**partial digest** A restriction digest that has not been allowed to go to completion, and thus contains pieces of DNA that have restriction endonuclease sites that have not been cleaved.

**partial dominance** *See* incomplete dominance.

**Pascal's triangle** A triangular array of numbers made up of the coefficients of the binomial expansion.

**path diagram** A modified pedigree showing only the direct line of descent from common ancestors.

**pedigree** A representation of the ancestry of an individual or family; a family tree.

**penetrance** The normal appearance of genetically controlled traits in the phenotype.

**peptidyl transferase** The enzymatic center responsible for peptide bond formation during translation at the ribosome.

**pericentric inversion** A chromosomal inversion that includes the centromere.

**permissive temperature** A temperature at which temperature-sensitive mutants are normal.

**PEST hypothesis** Degradation of a protein in less than two hours is signaled by a region within the protein rich in proline (P), glutamic acid (E), serine (S), or threonine (T).

**petite mutations** Mutations of yeast that produce small colonies, like those grown under anaerobic conditions.

**phages** *See* bacteriophages.

**phenocopy** A phenotype that is not genetically controlled but looks like a genetically controlled phenotype.

**phenotype** The observable attributes of an organism.

**pheromone** A chemical signal, analogous to a hormone, that passes information between individuals.

**phosphodiester bond** A diester bond linking two nucleotides together (between phosphoric acid and sugars) to form the nucleotide polymers DNA and RNA.

**photocrosslinking** A technique used to determine which moieties (proteins, DNA) are in close proximity during a particular process.

**photoreactivation** The process whereby dimerized pyrimidines (usually thymines) in DNA are restored by an enzyme (deoxyribodipyrimidine photolyase) that requires light energy.

**phyletic evolution** *See* anagenesis.

**phyletic gradualism** The process of gradual evolutionary change over time.

**phylogenetic tree** A diagram showing evolutionary lineages of organisms.

**physical map** Chromosomal map in which distances are in physical units of base pairs. These maps can be of microsatellite markers or of sequence-tagged sites.

**pili (fimbriae)** Hairlike projections on the surface of bacteria; Latin for "hair."

**plaques** Clear areas on a bacterial lawn caused by cell lysis due to viral attack.

**plasmid** An autonomous, self-replicating genetic particle, usually of double-stranded DNA.

**plastid** A chloroplast prior to the development of chlorophyll.

**pleiotropy** The phenomenon whereby a single mutant affects several apparently unrelated aspects of the phenotype.

**point centromere** The type of centromere, such as that found in *Saccharomyces cerevisiae,* that has defined sequences large enough to accommodate one spindle microtubule.

**point mutations** Small mutations that consist of a replacement, addition, or deletion of one or a few bases.

**polar bodies** The small cells that are the by-products of meiosis in female animals. One functional ovum and as many as three polar bodies result from meiosis of each primary oocyte.

**polarity** Meaning "directionality" and referring either to an effect seen in only one direction from a point of origin or to the fact that linear entities (such as a single strand of DNA) have ends that differ from each other.

**polar mutant** An organism with a mutation, usually within an operon, that prevents the expression of genes distal to itself.

**pollen grain** The male gametophyte in higher plants.

**poly-A tail** A sequence of adenosine nucleotides added to the $3'$ end of eukaryotic messenger RNAs.

**polycistronic** Referring to prokaryotic messenger RNAs that contain several genes within the same messenger RNA transcript.

**polygenic inheritance** *See* quantitative inheritance.

**polymerase chain reaction (PCR)** A method to amplify DNA segments rapidly in temperature-controlled cycles of denaturation, primer binding, and replication.

**polymerase cycling** The process by which a DNA polymerase III enzyme completes an Okazaki fragment, releases it, and begins synthesis of the next Okazaki fragment.

**polymerized** Formed into a complex compound by linking together smaller elements.

**polynucleotide phosphorylase** An enzyme that can polymerize diphosphate nucleotides without the need for a primer. The function of this enzyme *in vivo* is probably in its reverse role as an RNA exonuclease.

**polyploids** Organisms with greater than two chromosome sets.

**polyribosome** *See* polysome.

**polysome** The configuration of several ribosomes simultaneously translating the same messenger RNA. Shortened form of the term *polyribosome*.

**polytene chromosome** Large chromosome, seen, for example, in *Drosophila* salivary glands, consisting of many chromatids formed by rounds of endomitosis. Synapsis of homologous chromosomes occurs during the process.

**population** A group of organisms of the same species relatively isolated from other groups of the same species. *See* deme.

**position effect** An alteration of phenotype caused by a change in the relative arrangement of the genetic material.

**positive interference** When the occurrence of one crossover reduces the probability that a second will occur in the same region.

**postreplicative repair** A DNA repair process initiated when DNA polymerase bypasses a damaged area.

**posttranscriptional modifications** Changes in eukaryotic messenger RNA made after transcription has been completed. These changes include additions of caps and tails and removal of introns.

**preemptor stem** A configuration of leader transcript messenger RNA that does not terminate transcription in the attenuator-controlled amino acid operons.

**pre-initiation complex (PIC)** The form of the RNA polymerase II enzyme with general transcription factors bound equivalent to the *E. coli* holoenzyme. Phosphorylation of the enzyme then allows transcription to begin.

**Pribnow box** Consensus sequence of TATAAT in prokaryotic promoters centered at the position $-10$.

**primary oocytes** The cells that undergo meiosis in female animals.

**primary spermatocytes** The cells that undergo meiosis in male animals.

**primary structure** The sequence of polymerized amino acids in a protein.

**primary transcript** The product of eukaryotic transcription before posttranscriptional modification takes place.

**primase** An enzyme that creates a messenger RNA primer for Okazaki fragment initiation.

**primer** In DNA replication, a length of double-stranded DNA that continues as a single-stranded template in the $3'$ to $5'$ direction.

**primosome** A complex of two proteins, a primase and helicase, that initiates RNA primers on the lagging DNA strand during DNA replication.

**prion** Infectious agent responsible for several neurological diseases (scrapie, kuru, Creutzfeld-Jakob syndrome, mad-cow disease). It is a protein that lacks DNA or RNA.

**probability** The expectation of the occurrence of a particular event.

**probability theory** The conceptual framework concerned with quantification of probabilities. *See* probability.

**proband** *See* propositus.

**probe** In recombinant DNA work, a radioactive nucleic acid complementary to a region being searched for in a restriction digest or genomic library.

**processivity** The ability of an enzyme to repetitively continue its catalytic function without dissociating from its substrate.

**product rule** The rule that states that the probability that two independent events will both occur is the product of their separate probabilities.

**progeny testing** Breeding of offspring to determine their genotypes and those of their parents.

**prokaryotes** Organisms that lack true nuclei.

**promoter** A DNA region that RNA polymerase binds to in order to initiate transcription.

**proofread** Technically, to read for the purpose of detecting errors for later correction. DNA polymerase has 3′ to 5′ exonuclease activity, which it uses during polymerization to remove incorrect nucleotides it has recently added.

**prophage** A temperate phage replicating with the host that can later initiate the lytic cycle.

**prophase** The initial stage of mitosis or meiosis in which chromosomes become visible and the spindle apparatus forms.

**proplastid** Mutant plastids that do not grow and develop into chloroplasts.

**propositus (proposita)** The person through whom a pedigree was discovered.

**proteasome** A barrel-shaped cellular organelle for protein breakdown involving the ubuiquitin pathway.

**proteome** From *prote*ins of the gen*ome;* the complete set of proteins from a particular genome. It is the protein analogue to "genome."

**proteomics** The study of the complete set of proteins from a particular genome. It is the protein analogue to "genomics."

**proto-oncogene** A cellular oncogene in an untransformed cell.

**prototrophs** Strains of organisms that can survive on minimal medium.

**pseudoalleles** Genes that are functionally but not structurally allelic. Within gene families, pseudoalleles are alleles that are not expressed.

**pseudoautosomal gene** A gene that occurs on both sex-determining heteromorphic chromosomes.

**pseudodominance** The phenomenon in which a recessive allele shows itself in the phenotype when only one copy of the allele is present, as in hemizygous alleles or in deletion heterozygotes.

**P (peptidyl) site** The site on the ribosome occupied by the peptidyl-tRNA just before peptide bond formation.

**punctuated equilibrium** The evolutionary process involving long periods without change (stasis) punctuated by short periods of rapid speciation.

**Punnett square** A diagrammatic representation of a particular cross used to predict the progeny of the cross.

**purines** Nitrogenous bases of which guanine and adenine are found in DNA and RNA.

**pyrimidines** Nitrogenous bases of which thymine is found in DNA, uracil in RNA, and cytosine in both.

**quantitative inheritance** The mechanism of genetic control of traits showing continuous variation.

**quantitative trait loci** Chromosomal regions contributing to the inheritance of a quantitative trait. These regions may contain one or more polygenes that contribute to the phenotype.

**quantitative variation** *See* continuous variation.

**quaternary structure** The association of polypeptide subunits to form the final structure of a protein.

**random genetic drift** Changes in allelic frequency due to sampling error.

**random mating** The mating of individuals in a population such that the union of individuals with the trait under study occurs according to the product rule of probability.

**random strand analysis** Mapping studies in organisms that do not keep all the products of meiosis together.

**read-through** Transcription or translation beyond the normal termination signals in DNA or RNA, respectively.

**realized heritability** Heritability measured by a response to selection.

**recessive** An allele (or phenotype) that does not express itself in the heterozygous condition.

**reciprocal cross** A cross with the phenotype of each sex reversed as compared with the original cross. Made to test the role of parental sex on inheritance pattern.

**reciprocal translocation** A chromosomal configuration in which the ends of two nonhomologous chromosomes are broken off and become attached to the nonhomologues.

**recombinant DNA technology** Techniques of gene cloning. Recombinant DNA refers to the hybrid of foreign and vector DNA. *See* gene cloning.

**recombinant plasmid** A plasmid that contains an inserted piece of foreign DNA.

**recombinants** In mapping studies, offspring with allelic arrangements made up of a combination of the original parental alleles.

**recombination** The nonparental arrangement of alleles in progeny that can result from either independent assortment or crossing over.

**recombination nodule** Proteinaceous nodules found on bivalents during zygonema and pachynema associated with crossing over.

**recon** A term Benzer coined for the smallest recombinable unit within a cistron.

**reductional division** The first meiotic division. It reduces the number of chromosomes and centromeres to half that in the original cell.

**regional centromere** The type of centromere found in higher eukaryotes that can accommodate several spindle microtubules.

**regulator gene** A gene primarily involved in control of the production of another gene's product.

**reinitiation** The initiation of translation by a ribosome that has just completed translation of a region of the messenger RNA upstream of the current point of initiation.

**relative Darwinian fitness** *See* fitness.

**relaxed mutant** A mutant that does not exhibit the stringent response under amino acid starvation.

**release factors (RFl and RF2)** Proteins in prokaryotes responsible for termination of translation and release of the newly synthesized polypeptide when a nonsense codon appears in the A site of the ribosome. Replaced by eRF in eukaryotes.

**repetitive DNA** DNA made up of copies of the same nucleotide sequence.

**replica-plating** A technique to rapidly transfer microorganism colonies to numerous petri plates.

**replication-coupling assembly factor** A protein complex in fruit flies that assembles new nucleosomes.

**replicons** A replicating genetic unit including a length of DNA and its site for the initiation of replication.

**replisome** The DNA-replicating structure at the Y-junction, consisting of two DNA polymerase III enzymes and a primosome (primase and DNA helicase).

**reporter systems** Genetic constructs that allow an investigator to determine that a specific locus is active by measuring the phenotypic expression of an associated locus, such as the luciferase reporter, which glows if watered with luciferin.

**repressible system** A coordinated group of enzymes involved in a synthetic pathway (anabolic) is repressible if excess quantities of the end product of the pathway lead to the termination of transcription of the genes for the enzymes. These systems are primarily prokaryotic operons.

**repressor** The protein product of a regulator gene that acts to control transcription of inducible and repressible operons.

**reproductive isolating mechanisms** Environmental, behavioral, mechanical, and physiological barriers that prevent two individuals of different populations from producing viable progeny.

**reproductive success** The relative production of offspring by a particular genotype.

**repulsion** Allelic arrangement in which each homologous chromosome has mutant and wild-type alleles.

**resistance transfer factor** Infectious transfer part of R plasmids.

**restricted transduction** *See* specialized transduction.

**restriction digest** The results of the action of a restriction endonuclease on a DNA sample.

**restriction endonucleases** Endonucleases that recognize certain DNA sequences, then cleave them. They protect cells from viral infection; they are useful in recombinant DNA work.

**restriction fragment length polymorphism (RFLP)** Variations (among individuals) in banding patterns of electrophoresed restriction digests.

**restriction map** A physical map of a piece of DNA showing recognition sites of specific restriction endonucleases separated by lengths marked in numbers of bases.

**restriction site** The sequence of DNA recognized by a restriction endonuclease.

**restrictive temperature** A temperature at which temperature-sensitive mutants display the mutant phenotype.

**retinoblastoma** A childhood cancer of retinoblast cells caused by the inactivation of an anti-oncogene.

**retrotransposons** Transposable genetic elements found in eukaryotic DNA that move through the reverse transcription of an RNA intermediate.

**reverse transcriptase** An enzyme that can use RNA as a template to synthesize DNA.

**reversion** The return of a mutant to the wild-type phenotype by way of a second mutational event.

**R factors** *See* R plasmids.

**rho-dependent terminator** A DNA sequence signaling the termination of transcription; termination requires the presence of the rho protein.

**rho-independent terminator** A DNA sequence signaling the termination of transcription; the rho protein is not required for termination.

**rho protein** A protein involved in the termination of transcription.

**ribosomal RNA (rRNA)** RNA components of the subunits of the ribosomes.

**ribosome recycling factor (RRF)** A protein needed to prepare ribosomal subunits that have just finished translating a messenger RNA for another cycle of translation.

**ribosomes** Organelles at which translation takes place. They are made up of two subunits consisting of RNA and proteins.

**ribozyme** Catalytic or autocatalytic RNA.

**RNA editing** The insertion of uridines into messenger RNAs after transcription is completed; controlled by guide RNA. May also involve insertion of cytidines in some organisms or possible deletions of bases.

**RNA phages** Phages whose genetic material is RNA. They are the simplest phages known.

**RNA polymerase** The enzyme that polymerizes RNA by using DNA as a template. (Also known as *transcriptase* or *RNA transcriptase.*)

**RNA replicase** A polymerase enzyme that catalyzes the self-replication of single-stranded RNA.

**Robertsonian fusion** Fusion of two acrocentric chromosomes at the centromere.

**rolling-circle replication** A model of DNA replication that accounts for a circular DNA molecule producing linear daughter double helices.

**R plasmids** Plasmids that carry genes that control resistance to various drugs.

**rRNA** *See* ribosomal RNA (rRNA).

**rule of independent assortment** *See* independent assortment, rule of.

**rule of segregation** *See* segregation, rule of.

**sampling distribution** The distribution of frequencies with which various possible events could occur, or a probability distribution defined by a particular mathematical expression.

**sarcoma** Tumor of tissue of mesodermal origin (e.g., muscle, bone, cartilage).

**satellite DNA** Highly repetitive eukaryotic DNA primarily located around centromeres. Satellite DNA usually has a different buoyant density than the rest of the cell's DNA.

**scaffold** The eukaryotic chromosomal structure that remains when DNA and histones have been removed.

**scanning hypothesis** Proposed mechanism by which the eukaryotic ribosome recognizes the initiation region of a messenger RNA after binding the 5′ capped end of it. The ribosome scans the messenger RNA for the initiation codon.

**scientific method** A procedure scientists use to test hypotheses, making predictions about the outcome of an observation or experiment before the experiment is performed. The results provide support or refutation of the hypothesis.

**screening technique** A technique to isolate a specific genotype or phenotype of an organism.

**secondary oocytes** The cells formed by meiosis I in female animals.

**secondary spermatoctyes** The products of the first meiotic division in male animals.

**secondary structure** The flat or helical configuration of the polypeptide backbone of a protein.

**second-division segregation (SDS)** The allelic arrangement in the spores of Ascomycete fungi with ordered spores that indicates a crossover between a locus and its centromere

**securin** An inhibitory protein that prevents separin from acting on cohesin to separate sister chromatids.

**segmentation genes** Genes of developing embryos that determine the number and fate of segments.

**segregation, rule of** Mendel's first principle, which describes how genes pass from one generation to the next.

**segregational load** Genetic load caused when a population is segregating less fit homozygotes because of heterozygote advantage.

**segregation distortion** *See* gametic selection.

**selection** *See* natural selection.

**selection coefficients, *s, t*** The sum of forces acting to lower the relative reproductive success of a genotype.

**selection-mutation equilibrium** An equilibrium allelic frequency resulting from the balance between selection against an allele and mutation re-creating this allele.

**selective medium** A culture medium enriched with a particular substance to allow the growth of particular strains of organisms.

**selfed** *See* self-fertilization.

**self-fertilization** Fertilization in which the two gametes come from the same individual.

**selfish DNA** A segment of the genome with no apparent function, although it can control its own copy number.

**semiconservative replication** The mode by which DNA replicates. Each strand acts as a template for a new double helix. *See* template.

**semisterility** Nonviability of a proportion of gametes or zygotes.

**sense strand** *See* coding strand.

**separin** An enzyme that breaks down cohesin and allows sister chromatids to separate at the start of anaphase of mitosis.

**sequence-tagged sites (STSs)** DNA lengths of 100–500 base pairs that are unique in the genome. They are created by polymerase chain reaction amplification of primers that are then tested to be sure the sequence is unique.

**sex chromosomes** Heteromorphic chromosomes whose distribution in a zygote determines the sex of the organism.

**sex-conditioned traits** Traits that appear more often in one sex than in another.

**sex-determining region Y (*SRY*)** The sex switch, or testis-determining factor, in human beings, located on the Y chromosome (*Sry* in mice).

**sexduction** A process whereby a bacterium gains access to and incorporates foreign DNA brought in by a modified F factor during conjugation.

**sex-influenced traits** *See* sex-conditioned traits.

**Sex-lethal** A gene in *Drosophila*, located on the X chromosome, that is a sex switch, directing development toward femaleness when in the "on" state. It is regulated by numerator and denominator elements that act to influence the genic balance ratio (X/A).

**sex-limited traits** Traits expressed in only one sex. They may be controlled by sex-linked or autosomal loci.

**sex-linked** The inheritance pattern of loci located on the sex chromosomes (usually the X chromosome in XY species); also refers to the loci themselves.

**sex-ratio phenotype** A trait in *Drosophila* whereby females produce mostly, if not only, daughters.

**sex switch** A gene in mammals, normally found on the Y chromosome, that directs the indeterminate gonads towards development as testes.

**sexual selection** The forces, determined by mate choice, that act to cause one genotype to mate more frequently than another genotype.

**Shine-Dalgarno hypothesis** A proposal that prokaryotic messenger RNA is aligned at the ribosome by complementarity between the messenger RNA upstream from the Initiation codon and the 3′ end of the 16S ribosomal RNA.

**shoot apical meristem** The major meristematic tissue of the plant; surrounds the shoot.

**short interspersed elements (SINEs)** Sequences of DNA interspersed in eukaryotic chromosomes in many copies. Alu, a three-hundred base-pair sequence, is found about half a million times in human DNA.

**shotgun cloning** The random cloning of pieces of the DNA of an organism without regard to the genes or sequences present in the cloned DNA.

**shunting** Process in which the first initiation codon on a messenger RNA is bypassed for an initiation codon further down the messenger. The process is probably guided by secondary structure in the messenger.

**siblings (sibs)** Brothers and sisters.

**sigma factor** The protein that gives promoter-recognition specificity to the RNA polymerase core enzyme of bacteria.

**signal hypothesis** The major mechanism whereby proteins that must insert into or across a membrane are synthesized by a membrane-bound ribosome. The first thirteen to thirty-six amino acids synthesized, termed a *signal peptide*, are recognized by a *signal recognition particle* that draws the ribosome to the membrane surface. The signal peptide may be removed later from the protein.

**signal peptide** *See* signal hypothesis.

**signal recognition particle** *See* signal hypothesis.

**signal transduction pathway** A pathway in which the action of kinase enzymes that free transcription factors, or some other action, translates an environmental signal into some form of gene action.

**single-nucleotide polymorphisms (SNPs)** Differences between individuals involving single base pairs that are located about every 1,000 bases along the human genome. SNPs are useful for mapping disease genes.

**single-strand binding proteins** Proteins that attach to single-stranded DNA, usually near the replicating Y-junction, to stabilize the single strands.

**sister chromatids** *See* chromatids.

**site-specific recombination** A crossover event, such as the integration of phage λ, that requires homology of only a very short region and uses an enzyme specific for that recombination.

**small nuclear ribonucleoproteins (snRNPs)** Components of the spliceosome, the intron-removing apparatus in eukaryotic nuclei. *See* snRNPs.

**small nucleolar ribonucleoprotein particles (snoRNPs)** Particles composed of RNA and protein found in the nucleolus that modify ribosomal RNAs, particularly by converting some uridines to pseudouridines and methylating some ribose sugars.

**small nucleolar RNAs (snoRNAs)** RNAs found in small nucleolar ribonucleoprotein particles (snoRNPs) that take part in modifying ribosomal RNA in the nucleolus.

**SMC proteins** For *s*tructural *m*aintenance of *c*hromosomes; proteins that aid mitotic segregation, sister-chromatid adhesion, dosage compensation, recombination, and other chromosomal activities.

**snRNPs** *See* small nuclear ribonucleoproteins.

**sociobiology** The study of the evolution of social behavior in animals.

**somatic doubling** A disruption of the mitotic process that produces a cell with twice the normal chromosome number.

**somatic hypermutation** The occurrence of a high level of mutation in the variable regions of immunoglobulin genes.

**SOS box** The region of the promoters of various genes that the LexA repressor recognizes. Release of repression results in the induction of the SOS response.

**SOS response** Repair systems (RecA, Uvr) induced by the presence of single-stranded DNA that usually occurs from postreplicative gaps caused by various types of DNA damage. The RecA protein, stimulated by single-stranded DNA, is involved in the inactivation of the LexA repressor, thereby inducing the response.

**Southern blotting** A method, first devised by E. M. Southern, used to transfer DNA fragments from an agarose gel to a nitrocellulose gel for the purpose of DNA-DNA or DNA-RNA hybridization during recombinant DNA work.

**specialized transduction** Form of transduction based on faulty looping out by a temperate phage. Only neighboring loci to the attachment site can be transduced. *See* generalized transduction.

**speciation** A process whereby, over time, one species evolves into a different species (anagenesis) or one species diverges to become two or more species (cladogenesis).

**species** A group of organisms capable of interbreeding to produce fertile offspring.

**specific transcription factors** Proteins needed for activation of transcription at specific eukaryotic promoters. Also, negative factors that can inhibit transcription at a specific eukaryotic promoter.

**spermatids** The four products of meiosis in males that develop into sperm.

**spermatogenesis** The process of sperm production.

**spermatogonium** A cell type in the testes of male vertebrates that gives rise to primary spermatocytes by mitosis.

**sperm cells** The gametes of males.

**spermiogenesis** The process by which spermatids mature into sperm cells.

**spindle** The microtubule apparatus that controls chromosomal movement during mitosis and meiosis.

**spindle pole body** Spindle microtubule organizing center found in fungi.

**spiral cleavage** The cleavage process in mollusks and some other invertebrates whereby the spindle is tipped at mitosis in relation to the original egg axis.

**spirillum** A spiral bacterium.

**spliceosome** Protein-RNA complex that removes introns in eukaryotic nuclear RNAs.

**sporophyte** The stage of a plant life cycle that produces spores by meiosis and alternates with the gametophyte stage.

**stabilizing selection** A type of selection that removes individuals from both ends of a phenotypic distribution, thus maintaining the same distributional mean.

**standard deviation** The square root of the variance.

**standard error of the mean** The standard deviation divided by the square root of the sample size. It is the standard deviation of a sample of means.

**statistics** Measurements of attributes of a sample from a population; denoted by Roman letters. *See* parameters.

**stem-loop structure** A lollipop-shaped structure formed when a single-stranded nucleic acid molecule loops back on itself to form a complementary double helix (stem), topped by a loop.

**stepladder gel** A DNA-sequencing gel. The numerous bands in each lane give the appearance of a stepladder.

**stochastic** A process with an indeterminate or random element as opposed to a deterministic process with no random element.

**stringent factor** A protein that catalyzes the formation of an unusual nucleotide (guanosine tetraphosphate) during the stringent response under amino acid starvation.

**stringent response** A translational control mechanism in prokaryotes that represses transfer RNA and ribosomal RNA synthesis during amino acid starvation.

**structural alleles** Alleles whose alterations include some of the same base pairs.

**submetacentric chromosome** A chromosome whose centromere lies between its middle and its end, but closer to the middle.

**subtelocentric chromosome** A chromosome whose centromere lies between its middle and its end, but closer to the end.

**sum rule** The rule that states that the probability that two mutually exclusive events will occur is the sum of the probabilities of the individual events.

**supercoiling** Negative or positive coiling of double-stranded DNA that differs from the relaxed state.

**supergenes** Several loci, which usually control related aspects of the phenotype, in close physical association.

**suppressor gene** A gene that, when mutated, apparently restores the wild-type phenotype to a mutant of another locus.

**surveillance mechanism** Used to describe mechanisms that oversee checkpoints in the cell cycle, points in which the process can be halted if certain conditions are not met.

**survival of the fittest** In evolutionary theory, survival of only those organisms best able to obtain and utilize resources (the fittest). This phenomenon is the cornerstone of Darwin's theory of evolution.

**Svedberg unit** A unit of sedimentation during centrifugation. Abbreviation is S, as in 50S.

**swivelase** *See* DNA gyrase.

**sympatric speciation** Speciation in which reproductive isolating mechanisms evolve within the range and habitat of the parent species. This type of speciation may be common in parasites.

**synapsis** The point-by-point pairing of homologous chromosomes during zygonema or in certain dipteran tissues that undergo endomitosis.

**synaptonemal complex** A proteinaceous complex that apparently mediates synapsis during the zygotene stage and then disintegrates.

**syncitium** A cell that has many nuclei not separated by cell membranes.

**synexpression group** A group of eukaryotic genes that are involved in the same function or pathway and are induced together.

**synteny test** A test that determines whether two loci belong to the same linkage group by observing concordance in hybrid cell lines.

**synthetic medium** A chemically defined substrate that microorganisms are grown upon.

**TACTAAC box** A consensus sequence surrounding the lariat branch point of eukaryotic messenger RNA introns.

**TATA-binding protein** A protein, part of TFIID, that binds the TATA consensus sequence in eukaryotic promoters.

**TATA box** An invariant DNA sequence at about $-25$ in the promoter region of eukaryotic genes; analogous to the Pribnow box in prokaryotes.

**tautomeric shift** Reversible shifts of proton positions in a molecule. The bases in nucleic acids shift between the keto and enol forms or between the amino and imino forms.

**TBP-associated factors (TAFs)** Proteins that bind with the TATA-binding protein to form TFIID. They aid in the selectivity of TFIID.

**T-cell receptors** Surface proteins of T cells that allow the T cells to recognize host cells that have been infected.

**telocentric chromosome** A chromosome whose centromere lies at one of its ends.

**telomerase** An enzyme that adds telomeric sequences to the ends of eukaryotic chromosomes.

**telomeres** The ends of linear chromosomes.

**telophase** The terminal stage of mitosis or meiosis in which chromosomes uncoil, the spindle breaks down, and cytokinesis usually occurs.

**telson** The posterior end of the arthropod embryo, where the end of the alimentary canal is located.

**temperate phage** A phage that can enter into lysogeny with its host.

**temperature-sensitive mutant** An organism with an allele that is normal at a permissive temperature but mutant at a restrictive temperature.

**template** A pattern serving as a mechanical guide. In DNA replication, each strand of the duplex acts as a template for the synthesis of a new double helix.

**template strand** *See* anticoding strand.

**terminator sequence** A sequence in DNA that signals the termination of transcription to RNA polymerase.

**terminator stem** A configuration of the leader transcript that signals transcriptional termination in attenuator-controlled amino acid operons.

**tertiary structure** The further folding of a protein, bringing α helices and β sheets into three-dimensional arrangements.

**testcross** The cross of an organism with a homozygous recessive organism.

**testing of hypotheses** The determination of whether to reject or fail to reject a proposed hypothesis based on the likelihood of the experimental results.

**testis-determining factor (*TDF*)** General term for the gene determining maleness in human beings (*Tdf* in mice).

**tetrads** The meiotic configuration of four chromatids first seen in pachynema. There is one tetrad (bivalent) per homologous pair of chromosomes.

**tetranucleotide hypothesis** The hypothesis, based on incorrect information, that DNA could not be the genetic material because its structure was too simple—that is, that repeating subunits contain one copy each of the four DNA nucleotides.

**tetraploids** Organisms with four whole sets of chromosomes.

**tetratype (TT)** A spore arrangement in Ascomycete fungi that consists of the two parental and two recombinant spores.

**theta structure** An intermediate structure formed during the replication of a circular DNA molecule.

**three-point cross** A cross involving three loci.

**thymine** *See* pyrimidines.

**t-loop** A loop that forms at the end of mammalian telomeres by the interdigitation of the $3'$ free end into the DNA double helix.

**topoisomerase** An enzyme that can relieve (or create) supercoiling in DNA by creating transitory breaks in one (type I) or both (type II) strands of the helical backbone.

**topoisomers** Forms of DNA with the same sequence but differing in their linkage number (coiling).

**totipotent** The state of a cell that can give rise to any and all adult cell types, as compared with a differentiated cell whose fate is determined.

**trailer** The length of messenger RNA from the nonsense codon to the $3'$ end or, in polycistronic messenger RNAs, from a nonsense codon to the next gene's leader.

***trans*** Meaning "across" and referring usually to the geometric configuration of mutant alleles across from each other on a homologous pair of chromosomes.

***trans*-acting** Referring to mutations of, for example, a repressor gene, that act through a diffusible protein product; the normal mode of action of most recessive mutations.

**transcription** The process whereby RNA is synthesized from a DNA template.

**transcription factors** Eukaryotic proteins that aid RNA polymerase in recognizing promoters. *See* general transcription factors *and* specific transcription factors.

**transducing particle** A defective phage, carrying part of the host's genome.

**transduction** A process whereby a cell can gain access to and incorporate foreign DNA brought in by a viral particle.

**transfection** The introduction of foreign DNA into eukaryotic cells.

**transfer operon (*tra*)** Sequence of loci that impart the male (F-pili-producing) phenotype on a bacterium. The male cell can transfer the F plasmid to an $F^-$ cell.

**transfer RNA (tRNA)** Small RNA molecules that carry amino acids to the ribosome for polymerization.

**transformation** A process whereby prokaryotes take up DNA from the environment and incorporate it into their genomes, or the conversion of a eukaryotic cell into a cancerous one.

**transgenic** Eukaryotic organisms that have taken up foreign DNA.

**transition mutation** A mutation in which a purine-pyrimidine base pair replaces a base pair in the same purine-pyrimidine relationship.

**translation** The process of protein synthesis wherein the nucleotide sequence in messenger RNA determines the primary structure of the protein.

**translocase (EF-G)** Elongation factor in prokaryotes necessary for proper translocation at the ribosome during the translation process. Replaced by eEF2 in eukaryotes.

**translocation** A chromosomal configuration in which part of a chromosome becomes attached to a different chromosome. Also a part of the translation process in which the messenger RNA is shifted one codon in relation to the ribosome.

**translocation channel (translocon)** A protein-lined pore or channel in a membrane through which nascent proteins are transported during translation.

**transposable genetic element** A region of the genome, flanked by inverted repeats, a copy of which can be inserted at another place; also called a transposon or a jumping gene.

**transposon** *See* transposable genetic element.

**transversion mutation** A mutation in which a purine replaces a pyrimidine, or vice versa.

**trihybrid** An organism heterozygous at three loci.

**triploids** Organisms with three whole sets of chromosomes.

**trisomic** A diploid cell with an extra chromosome.

**tRNA** *See* transfer RNA (tRNA).

***trp* RNA-binding attenuation protein (TRAP)** Protein that can bind to the attenuation region of the messenger RNA of the tryptophan operon in *Bacillus subtilis,* causing a terminator stem to form and halting further transcription.

**true heritability** *See* heritability.

**tumor** Abnormal growth of tissue.

**tumor-suppressor genes** Genes that normally prevent unlimited cellular growth. When both copies of the gene are mutated, cellular transformation follows. Examples are the *p53* gene and the genes for retinoblastoma and Wilm's tumor.

**two-point cross** A cross involving two loci.

**type I error** In statistics, the rejection of a true hypothesis.

**type II error** In statistics, the accepting of a false hypothesis.

**typological thinking** The concept that organisms of a species conform to a specific norm. In this view, variation is considered abnormal.

**ubiquitin** A peptide of twenty-six amino acid residues that enzymes attach to proteins that the proteasome will degrade.

**unequal crossing over** Nonreciprocal crossing over caused by the mismatching of homologous chromosomes. Usually occurs in regions of tandem repeats.

**uninemic chromosome** A chromosome consisting of one DNA double helix.

**unique DNA** A length of DNA with no repetitive nucleotide sequences.

**unmixed families** Groups of four codons sharing their first two bases and coding for the same amino acid.

**unusual bases** Other bases, in addition to adenine, cytosine, guanine, and uracil, found primarily in transfer RNAs.

**UP element** *See* upstream element.

**upstream** A convention on DNA related to the position and direction of transcription by RNA polymerase ($5' \rightarrow 3'$). Downstream (or $3'$ to) is in the direction of transcription whereas upstream ($5'$ to) is in the direction from which the polymerase has come.

**upstream element** A sequence of about twenty AT-rich bases centered at −50 in promoters of prokaryotic genes that are expressed strongly.

**uracil** *See* pyrimidines.

**variable-number-of-tandem-repeats (VNTR) loci** Loci that are hypervariable because of tandem repeats. Presumably, variability is generated by unequal crossing over.

**variance** The average squared deviation about the mean of a set of data.

**variegation** Patchiness; a type of position effect that results when particular loci are contiguous with heterochromatin.

**virion** A virus particle.

**viroids** Bare RNA particles that are plant pathogens.

**V(D)J joining** The process of joining variable, diversity, and joining gene segments (V-J, V-D, or D-J joining) in the formation of a functioning immunoglobulin gene.

**Wahlund effect** The effect of subdivision on a population, causing it to contain fewer heterozygotes than predicted despite the fact that all subdivisions are in Hardy-Weinberg proportions.

**western blotting** A technique for probing for a particular protein using antibodies. *See* Southern blotting.

**whole-genome shotgun method** A method of sequencing entire genomes by breaking up the genomes into small pieces, sequencing the pieces, and then using computers to establish order by overlapping the sequences.

**wild-type** The phenotype of a particular organism when it is first seen in nature.

**Wilm's tumor** A childhood kidney cancer caused by the inactivation of an anti-oncogene.

**wobble** Referring to the reduced constraint over the third base of an anticodon as compared with the other bases, thus allowing additional complementary base pairings.

**xeroderma pigmentosum** A disease in human beings caused by a defect in the UV mutation repair system.

**X-inactivation center (*XIC*)** Locus at which inactivation is initiated on the X chromosome in mammals.

**X linked** *See* sex-linked.

**X-ray crystallography** A technique, using X rays, to determine the atomic structure of molecules that have been crystallized.

**yeast artificial chromosome (YAC)** Originating from a bacterial plasmid, a YAC contains additionally a yeast centromeric region (CEN) and a yeast origin of DNA replication (ARS). YACs are capable of including large pieces of foreign DNA during cloning.

**Y-junction** The point of active DNA replication where the double helix opens up so that each strand can serve as a template.

**Y linked** Inheritance pattern of loci located on the Y chromosome only. Also refers to the loci themselves.

**Z DNA** A left-handed form of DNA found under physiological conditions in short GC segments that are methylated. It may be involved in regulating gene expression in eukaryotes.

***ZFY* gene** Originally believed to be the human male sex-switch gene, located on the short arm of the Y chromosome. *ZFY* stands for *z*inc *f*inger on the *Y* chromosome.

**zinc finger** Configuration of a DNA-binding protein that resembles a finger with a base, usually cysteines and histidines, binding a zinc ion. Discovered in a transcription factor in *Xenopus.*

**zygonema (zygotene stage)** The stage of prophase I of meiosis in which synapsis occurs.

**zygotic induction** The beginning of vegetative growth when a prophage is passed into an $F^-$ cell during conjugation.

**zygotic selection** The forces acting to cause differential mortality of an organism at any stage (other than gametes) in its life cycle.

# INDEX

**A**

ABC excinuclease, 342
Abelson, J., 268
ABO blood groups, 25–26, *32*, 559
Ac-Ds system, 468
Acentric fragment, 178, *179*
*Acetabularia* (green alga), 515
*Acinonyx jubatus*. *See* Cheetah
Acquired immunodeficiency syndrome. *See* AIDS
Acridine dyes, *337*
Acrocentric chromosome, 48, *49*
Acron, 470–71
Activators, for transcription, 263
Active site, of enzyme, 206, *208*
Acute lymphocytic leukemia, *485*
*Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Williams, 1966), 604–5
Adaptive mutations, 339
Adaptive value, 577
  Additive models, of variation, 533, 538, 548
Adenines
  chemistry of nucleic acids, 213, *215*
  dideoxy method of DNA sequencing, 385
  tautomeric shifts, *329*, 330
Adenosine deaminase, 397
Adjacent-1 and adjacent-2 types, of segregation, 184–85
A DNA, 219
Affected individuals, and pedigree analysis, 98
Affymetrix, Inc., 396
Africa, and AIDS, 502
African sleeping sickness, 275
AGAMOUS gene, 483
Agriculture
  genetically modified crops, 371, 398
  life cycle of plants, 66
  polyploidy in plants and, 199
  recombinant DNA techniques and, 13
*Agrobacterium tumefaciens*, 371–72, 376–77
AIDS, 13, 276, 502–4
Alberts, B., 234, 252
Albinism, 33, 34, 37, 97–98
Alcoholism, 103–4
Alkaptonuria, 37, 38, 582
Allele. *See also* Allelic frequency; Allelism
  classical genetics and, 9
  dominant and recessive traits, 18
  Hardy-Weinberg equilibrium, 558–59
  lethal-equivalent, 561
  multiple, 25–26, 558–59
Allelic exclusion, 495
Allelic frequency. *See also* Allele
  Hardy-Weinberg equilibrium, 553–54, 555, 561–62, 566–67
  heterozygote advantage, 584
  mutational equilibrium, 572–73
  selection-mutation equilibrium, 581
Allelism, *319*, 354. *See also* Allele
*Allium cepa*. *See* Onion
Allolactose, *407*, 408, 409
Allopatric speciation, 592, *592*, 594
Allopolyploidy, 198
Allosteric protein, 409
Allotypes, 493
Allozygosity, 561
Alternate segregation, 184
Alternative energy, and biotechnology, 398
Alternative splicing, 271–72
Altman, Sidney, 265, *268*
Altruism, and sociobiology, 14, 603–5, 606, 607–8
Alu family, 459
American Red Cross, 26
Ames, Bruce, 332
Ames test, for carcinogens, 332
Amino acids
  experimental methods and sequencing of, 284–86
  mutations, 313
  proteins and structure of, *282*
  rate of evolutionary change and sequences of, 601–602, *603*
  RNA and, *12*, 281
  sequence of dihydrofolate reductase, *513*
Aminoacyl site (A site), 292
Aminoacyl-tRNA synthetases, 257, 286, 287
*ampC* gene, 389
Amphidiploids, 198
Ampicillin, 427
Anagenesis, 590
Anaphase, of mitosis, 52, 54, *56*, *57*
Anaphase A and anaphase B, 54
Anaphase I, and meiosis, 59, *60*, *62*
Anaphase-promoting complex (APC), 51
Anemia. *See* Thalassemia
Aneuploids and aneuploidy, 83, 190, 192–97
Anfinsen, Christian, 303
Angelman syndrome, 524
Angiogenesis, 492
Angiosperms, 64
*Animal Dispersion in Relation to Social Behavior* (Wynne-Edwards, 1962), 603–4
Animals. *See also* Cattle; Mouse; Sheep
  life cycle, *47*
  meiosis, 63–64
  polyploidy, 198–99
  vectors and cloning, 371
*Antennapedia complex* (ANT-C), 478
Antibiotics and antibiotic resistance. *See also* Drug resistance
  antibacterial mechanisms, *153*
  *Chlamydomonas*, 517–18
  composite transposons, 427
  mitochondrial protein synthesis, 515
  resistance plasmids, 523
  translation and, 294–95
Antibodies
  ABO blood group system, 25
  definition of, 492
  diversity of, 494–97
  molecular genetics and, 13
Anticoding strand, of DNA, 249
Anticodons, 256, *258*, 309. *See also* Codons
Antigens, 25, 492
Antimutator mutations, 338
Anti-oncogenes, 13, 485
Antiparallel strands, of DNA, 218
Antirepressor, 421
*Antirrhinum majus* (Snapdragon), 35–36, 484
Antisense RNA, 431
Anti-sigma factor (AsiA), 430
Antiterminator protein, 420
AP endonuclease, 341, 354
Apolipoprotein-B (*apoB*) gene, 275
Apoptosis, 484
Apple maggot fly (*Rhagoletis pomonella*), 593
*Arabidopsis thaliana* (Meadow weed)
  dwarf form of, *373*
  flower formation, 482–83
  methylation of DNA, 466–67
  study of plant genetics and, 372
Arabinose, 435
Arberg, W., 4, 359
Archaea, 149
*Armillaria bulbosa* (fungus), 441
Aromatic amines, *492*
Arsenic, *492*
Asbestos, *492*
*Ascaris* spp., *50*, 87
Ascopores, 123
Ascus, 123
AsiA (anti-sigma factor), 430
Aspartate transcarbamylase, 433, *434*
*Aspergillus nidulans* (fungus), 132
Assignment test, 136–37
Assortative mating, 554, 560
Aster, 53
Asthma, *545*
Ataxia-telangiectasia, 484
Attenuator-controlled system, of gene expression, 415–18
Attenuator region, 415

Attenuator stem, 416
Autogamy, 519, 520, *521*
Autonomously replicating sequences (ARS), 239
Autopolyploidy, 198
Autoradiography, and DNA replication, 222–24
Autosomal dominant inheritance, 102, 104
Autosomal linkage, 134
Autosomal recessive inheritance, 102
Autosomal set, 84
Autosomes, 83
Autotrophs, 150
Autozygosity, 561, 564
Auxillary factors, and introns, 271
Auxotrophs, 150
Avery, Oswald, 4, 154, 209
AZT, 504

**B**

Bacillus, 149
*Bacillus subtilis* (soil bacterium)
  DNA-RNA complementarity, 247
  gene transcription, *253*
  transformation and transformation mapping, 155, 156–57, 172
  *trp* oepron, *417*, 418
*Bacillus thuringiensis* (Bt), 398
Bacitracin, *153*
Backcross, 22
Back mutations, 327
Bacteria. *See also Escherichia coli*; Prokaryotes
  antibiotic resistance, 294–95
  cultivation of, 150–51
  genetic research and, 149–50
  life cycle, 64
  phenotypes of, 151–54
  recombinant DNA technology, 349–51
  sexual processes in, 154–62
  size of cells, 441
Bacterial artificial chromosomes (BACs), 393, 395
Bacterial lawn, 151
Bacterial viruses, *66*, 149–50. *See also* Viruses
Bacteriophages
  genetic research and, 8, 149–50
  life cycle of, 163–65
  sexual processes in, 154–62
  transduction, 165–68
*Balaenoptera musculus* (Blue whale), 441
Balbiani, E. G., 449
Balbiani rings, 449
Baltimore, David, 276
*Bam*HI restriction endonuclease, 360
Barr, M., 90
Barr body, 90
*Bar* system, 185–86, *188*
Basal bodies, 52
Base excision repair, 340–41
Base flipping, 341
Basic/helix-loop-helix/leucine zipper, 480, *481*
Bates, H. W., 604
Batesian mimicry, 604, *605*
Bateson, William, 21
B-cell chronic lymphocytic leukemia, *485*
B DNA, 219, *220*
Beadle, George, 10, 38–39
*Beagle* (ship), 589
Bean (*Phaseolus vulgaris*), 541. *See also* Broad bean
Behavioral genetics, human, 547. *See also* Sociobiology
Bennett, J. C., 494
Benzer, Seymour, 318, 320, 321, 322–23, 324
Benzine, *492*
Berg, Paul, 4, 370, 383
Bertram, E., 90
*Bgl*I restriction enzyme, 360
*Bicoid* gene, 472, 474, 475
Binary fission, 519
Binomial expansion, 73
Binomial theorem, 72
Biochemical genetics, 37–39
Bioinformatics, 396
Biolistic transfer, 374, 375
Biological containment, 370–71
Biological species concept, 590
Biomedical applications
  AIDS and retroviruses, 502–4
  Ames test for carcinogens, 332
  antibiotics and antibiotic resistance, 294–95
  of multiple-stranded DNA, 221, *222*
  of prions, 213
Biometrics, 534–35
*Bithorax* gene, 478, *479*
Bivalents, and meiosis, 58
Blackburn, Elizabeth H., 454, 455
Blastoderm, 470
Blobel, Gunter, 301
Blood coagulating factor III, *136*
Blood pressure, and concordance, *546*
Blood transfusions, 26
Blood types
  Hardy-Weinberg equilibrium, 557, *558*, 559
  migration, 573–74
  phenotypes, 25–26, *32*
Blunt-end ligation, 363–64
Boa constrictor (*Constrictor constrictor*), *50*
*Bobbed* gene, 87, 96
Bodmer, W., 562
Body plan, 471–75
Bolivar, F., 363
*Bos taurus*. *See* Cattle
Bottlenecks, genetic, 576–77
*Bounty* (ship), 576
Bouquet stage, 56
Boveri, Theodor, 3
Boyer, H., 360
Branch migration, 348, *350*
*Brassica oleracea* (cabbage), 198
Breakage-fusion-bridge cycle, 178
Breakage-and-reunion process, 347
Breast cancer, 390, 391, *392*, *546*
Brenner, Sydney, 325, 483
Brewer, B., 252
Bridges, Calvin B., 84, 121, 199
Britten, Roy J., 457–58
Broad bean (*Vicia faba*), *50*
Brody, E., 268
Broker, Thomas, 265, *266*
Brower, L., 604
Brown, Robert, 3
Buoyant density, of DNA, 451, *452*, 514
Burkitt's lymphoma, *485*
Burt, Cyril, 546
Bush, Guy, 592
Butterflies, and mimicry, 604, *605*

**C**

Cabbage. *See Brassica oleracea*
Cactus ground-finch (*Geospiza scandens*), *588*, *595*
*Cactus* protein, *467*
*Caedobacter conjugatus* (kappa particle), 521
*Caedobacter taeniospiralis* (kappa particle), 520, *522*
*Caenorhabditis elegans* (Roundworm)
  genetic control of development, 483–84
  genome sequencing, 396
Cairns, John, 222–24, 339
Calculus of genes, 606
*Cameleo pardalis*. *See* Chameleon
Cancer. *See also* Breast cancer
  Ames test for carcinogens, 332
  environmental causes of, 492
  gene expression, 484
  gene mapping, 390, 391, *392*
  mutational nature of, 484–87, 506
  recombinant DNA technology, 13
  viral nature of, 487–92, 506
Cancer-family syndromes, 484–85
Cannabalism, and kuru, 213
Cap, and eukaryotic transcription, 265
Capsids, 149
Capsomeres, 149
Carcinogens
  Ames test, 332
  environmental causes of cancer, 492
Carcinomas, 484
Carpel whorl, 479, *482*
Cassette mechanism, 469
Cat, tortoiseshell, 91, *95*
Catabolite activator protein (CAP), 412–13
Catabolite repression, 412–13
*Catostomus clarki* (fish), 599, *600*
Cat's eye syndrome, 195
Cattle (*Bos taurus*), *66*, *543*
Cavalli-Sforza, L., 562
C-bands, 451, *452*
CCR5 protein, 502
*Cdc2* gene, 50–51
CD4 protein, 502
Cech, Thomas, 265–66, *268*
Celera Genomics, 396
Cell. *See also* Meiosis; Mitosis
  chromosomes and complements of, 48–49
  daughter cells, 49, *57*
  diploid cells, 9, 48
  disomic cells, *197*
  eukaryotic compared to prokaryotic, 440
  expression of foreign DNA and eukaryotic, 372–75

  F cells, 160, 161
  follicle cells, 470, *474*
  haploid cells, 9, 48
  helper T cells, 502
  mitosis and lymphocytic cells, *15*
  monosomic cells, 190
  nullisomic cells, 190
  nurse cells, 470
  pole cells, 470
  size of, 441
  sperm cells, 63
  totipotent cells, 469
  trisomic cells, 190
  upregulation of growth, 487
Cell cycle, 50–51
Cell-free system, 306
Cellular immunity, 492
Centimorgan, 111
Central dogma, of genetic information, 244, 275–76
Centric fragment, 178
Centrioles, 52–53
Centromeres, 48, 453, *454*
Centromeric breaks, 185
Centromeric fission, 185
Centrosome, 52, 53
Chain-terminating dideoxy nucleotides, 383–84
Chameleon (*Cameleo pardalis*), *2*
Chaperone proteins, 303
Chaperonins, 303
Chapeville, F., 287
Chargaff, Erwin, 216
Chargaff's ratios, 216
Charon phages, 360, 363
Chase, Martha, 209–10
Checkpoints, and cell cycle, 51
Cheetah (*Acinonyx jubatus*), *552*
Chemical method, of DNA sequencing, 383, 390
Chemical mutagenesis, 330–31, 336–37
Chemiluminescent techniques, 367
Chiasmata, 59
Chicken. *See also* Poultry
  epistatic interactions and phenotype of, *37*
  genotypic interactions and combs of, 30–32, *34*
  sex-linked traits, 96, *98*
Chimeras, 190
Chimeric plasmid, 360
Chirality, on amino acid, 281
*Chironomus pallidivittatus* (midge), *449*
*Chironomus tentans* (midge)
  chromosome puffs, 449, 450
  messenger RNA, *300*
Chi site, 350
Chi square analysis, 76–77, 79, 557, *558*, 567
*Chlamydomonas reinhardi* (green alga), 517–18
Chloramphenicol, 294, 515
Chloroplasts, and cytoplasmic inheritance, 515–18
Chondrodystrophy, 582
*Chorthippus parallelus*. *See* Grasshopper
Chow, Louise T., 265, *266*
Chromatids, 52, 67
Chromatin
  definition of, 48
  higher-order structure of, 448
  nucleosome structure, 443–44
  types of in eukaryotic chromosomes, *453*
Chromatin assembly factors, 444
Chromatin remodeling, 446, 465
Chromatosome, 443
Chromium, *492*
Chromomeres, 48
Chromosomal banding, 134, 451–53
Chromosomal breaks, 178–85, 200
Chromosomal maps. *See also* Mapping
  of *Drosophila melanogaster*, 121, *122*, *169–71*
  of humans, 132–40
Chromosomal painting, 486
Chromosomal rearrangements, in humans, 186–90
Chromosomal theory, 3–4, 66
Chromosome
  attachment point, 48
  cancer and defects in, 484, *485*
  chromatids compared to, 67
  combinations of maternal and paternal in gametes, 67
  complements of cell, 48–49
  cytogenetics and variations in structure of, 178–90
  cytological crossing over, 120–22
  definition of, 154
  Down syndrome, *177*
  eukaryotic, 440–61, 462
  sex determination, 83
  species and numbers of, 49–50
  three-point cross, 119
  two-point cross, 111, 114
Chromosome jumping, 390–91
Chromosome number, 190–99
Chromosome puffs, 449–50, 462
Chromosome walking, 390–91, *392*
Chronic myelogenous leukemia, *485*
Chronic wasting disease, 213
Chymotrypsin, 284
Cigarettes, and cancer, *492*
*Cis*-acting mutants, 411
*Cis* configuration, 111
*Cis*-dominant mutation, 411, 412
*Cis*-trans arrangement, and X linkage, 133
*Cis-trans* complementation test, 318
Cistron, 318
Cladogenesis, 590, 592–94, *595*
Clamp-loader complex, 232
Classical genetics, 4–5, 9–10, 66
Classical linkage maps, 393
Cleft lip, *545*
Clinal selection, 600
Clonal cancers, 484
Clonal evolution theory, of cancer, 484
Clones and cloning. *See also* Recombinant DNA technology
  assignment test, 137
  benefits from, 397–98
  chromosome location of gene, 141
  ethical debate on, 374–75
  particular genes, 364–66
  probing for cloned gene, 368–77
  restriction enzymes, 360–61
  of sheep, *1*
Clubfoot, and concordance, *546*
*C-myc* gene, 490
Coal products, as carcinogens, *492*
Coccus, 149
Coding strand, of DNA, 249
*Codium* (green alga), 515
Codominance, 23
Codon library, 602–3
Codon preference, 431
Codons. *See also* Anticodons
  common and alternative meanings of, *311*
  synthetic, 306–7
  transcription, 11–12
  transfer RNA, 256
  wobble hypothesis, 307–8, 309
Coefficient of coincidence, 118
Coefficient of relationship, 605–6
Cohen, S., 360
Cohesin, 53
Coiling, direction of in snails, 509–10, 525
Cointegrate state, of transposition, 428
Colicins, 522–23
Colinearity, and mutations, 324–25
Collins, Francis S., 358
Colony morphology
  bacteria, 151
  fungi, 124
Color. *See also* Pigment and pigmentation; Skin color
  flowers and inheritance of, 23, *37*
  metabolic pathways, *36*
  mouse and genotypes, 33–34
  multilocus control of in wheat, 531–33
  proplastids and variegation in plants, 516
Color blindness, 133
Col plasmids, 522–23
*Columba livia*. *See* Pigeon
Combinatorial control, 264–65
Comings, David E., 452
Common ancestry, 560–61
Comparative studies, 8
Compensasome, 94
Competence factor, 155
Complement, 492
Complementarity
  DNA-RNA hybridization, 246–47
  double helix of DNA, 11
Complementary DNA (cDNA), 365–66
Complementation, and mutations, 318–19
Complementation groups, 321
Complete linkage, 115
Complete medium, 150
Complementarity, and structure of DNA, 218, *219*
Complementation matrix, *320*, *323*
Component of fitness, 577
Component numbers, and transfer RNA, 281, 286
Composite transposons, 427
Computer programs, and population genetics, 583–84

Concordance, among twins, 546
Condensin, 444
Conditional-lethal mutant, 150, 327–28
Confidence limits, 74
Congenital malformations, and inbreeding, 562
Congenital heart disease, *545*
Conjugation, 154, 157, *158*, 161
Consanguineous degrees, of relationship, 99
Consensus sequence, 248
Conservative replication, of DNA, 220
Conserved sequence, 248, 277–78
Constitutive heterochromatin, 452–53
Constitutive mutants, 410–12
Constitutive puffs, 450
*Constrictor constrictor*. *See* Boa constrictor
Containment, and recombinant DNA technology, 370–71
Contigs, 393, *394*, *395*
Continuous replication, of DNA, 225
Continuous variation, 17, 531, 533
Control, and experimental design, 6
Controlling elements, and transposons, 468
Corepressor, 414
Corey, R. B., 206
Corn. *See Zea mays*
Correlation and correlation coefficient, 539–40
Correns, Carl, 4, 17
Cotransduction, 166, *167*
Coulson, Alan, 383
Coupling, of alleles, 111
Covariance, 539–40
CpG islands, 524
Creationism, 591
Creighton, Harriet, 120–22, *123*
Crick, Francis, 4, 10, 11, 205, 215, 216, 244, 272, 276, 304, 328, 330
Cricket (*Gryllus domesticus*), *50*
Cri du Chat syndrome, 188–90
Crisscross pattern of inheritance, 96
Critical chi-square, 76
*Cro*-gene product, 423
Crossbreeding, 17
Cross-fertilization, 17–18
Crossing over, and meiosis, 58–59
Crossover suppression, 179–80
Crow, James, 88, 533, 534, 562
Crown gall tumor, 371, *372*, 377
C-*sis* oncogene, 491
*Cucurbita pepo*. *See* Summer squash
Cultivation
  of bacteria, 150–51
  of viruses, 151
CURLY LEAF gene, 483
CUX family, of codons, 309
C-value paradox, 457
Cyanobacteria, and introns-early hypothesis of exon shuffling, 274
Cyclic AMP (cAMP), 412–13
Cyclin, 50–51
Cyclin-dependent kinase (CDK), 50
Cyclosome, 51
Cysteines, *152*, *283*
Cystine, *283*
Cystic fibrosis, 13
Cytogenetics. *See also* Molecular genetics
  chromosomal structure, 178–90
  chromosome number, 190–99
  definition of, 178
Cytokinesis, 47, 49–50
Cytological crossing over, 120–22, *123*
Cytoplasmic inheritance. *See also* Inheritance
  chloroplasts, 515–18
  definition of, 509
  infective particles, 518–22
  mitochondria, 511–15
  prokaryotic plasmids, 522–24
Cytosine, 213, *215*
Cytotoxic T lymphocytes, 492

## D

Damage reversal, and DNA repair, 340
*Danaus plexippus*. *See* Monarch butterfly
Danna, K., *359*
Darnell, J., 273
Darwin, Charles, 3, 4, 5, 13, 17, 589, 594
Dauermodification, 509
Daughter cells, 49, *57*
Davies, David, 221
DDT, resistance to in *Drosophila*, 533–34, *535*, 540
Dean, D., 252
Deer mouse (*Peromyscus maniculatus*), *92*
Degenerate code, 307
Degrees of freedom, 76, 79
Delbrück, Max, 4, 163, 164, 316, 317
*Deleted in colorectal cancer* (DCC) gene, 491
Deletion chromosome, 178, 179
Deletion mapping, 321–23
Deletions, and mutations, 336–37
Demes, 553
Denaturation studies, of DNA, 218
Denominator elements, 85
Density-gradient centrifugation, 221
Depauperate fauna, 594
Depression, and heritability, *545*
Derepressed operon, 414
Dervan, P., 221
Deterministic population size, 574
Development
  definition of, 469
  gene expression and patterns, 469–84
  hemoglobins and human, 459–61
Diabetes mellitus, *545*
Diakinesis, 56
Dicentric chromosome, 178, *179*
Dictyotene stage, of meiosis, 64
Dideoxyinosine, 504
Dideoxy method, of DNA sequencing, 383–86, 390
Dideoxynucleotides (dd), 384–85
Diener, Theodor, 272
Dihybrids, 26, 30, *33*, *36*
Dihydrofolate reductase, *513*
Dimerization, 340
Dioecious flowers, 87
Diploid cells, 9, 48
Diploid mapping, 110–11, 114–22
Diplonema, 56
Directed mutation, 339
Directional selection, 577–78
Disassortative mating, 560
Discontinuous DNA replication, 225, *227*
Discontinuous variation, 17, 531
Discrete generations, 556–57
Disomic cells, *197*
Dispersive replication, of DNA, 220
Disruptive selection, 578, *579*
Dissortative mating, 554
Disulfide bridge, and cysteines, *283*
Dizygotic twins, 546
D-loop model, of DNA replication, 238, *239*
DNA. *See also* Complementary DNA; Mitochondrial DNA; Mutation; Repetitive DNA; Satellite DNA; Transcription; Translation; Triplex DNA; Z DNA
  alternative forms of, 219
  biologically active structure, 214–15
  chemistry of, 211–19
  cloning and creation of, 364–66
  computer-generated images, *204*, *315*
  control of proteins, 205–7
  denaturation studies, 218
  double helix structure, 10–11, 206–7, 216–18
  enzymes and, 218–19
  eukaryotic cells and expression of foreign, 372–75
  eukaryotic chromosome and arrangement of, 440–42
  genomic library, 366
  location of, 208–9, 219
  methylation of, 466–67
  misalignment mutagenesis, *337*
  normal and tautomeric forms of bases, *328*, *329*
  partitioning of, 238
  phage labeling, 209–10
  plasmids, *357*
  protein motifs of recognition, 480–81
  RecA protein, *345*
  recombinant DNA technology, 12–13, 347–51, *352*
  repair, 339–46, *348*
  replication of, 207, 219, 220–39, 382
  translation, 276
  transcription, 246–55, 260–75
  transformation, *155*, 156, 209
  variation, 602–3
  X-ray crystallography, 215–16
DNA-DNA hybridization, 457–58
DNA fingerprints, 381, 382. *See also* Fingerprints
DNA glycosylases, 340–41
DNA gyrase, 236
DNA helicase, 230
DNA ligase, 227–29
DNA polymerase, 225, 227, *229*, *230*
DNA-RNA hybridization, 246–47
DNA sequencing, 383–90
DNA tumor viruses, 371
Docking protein (DP), 301
Dominance and dominant traits
  heterozygotes and homozygotes, 22–23

  - Mendel's experiments, 18, *19*
  - pedigree analysis, 99, 102
- Doolittle, W. E., 273
- *Dorsal* gene, 466, *467*
- Dosage compensation, 90–95
- Dot blotting, 368
- *Doublebar* gene, 185–86, *188*
- Double breaks, in chromosome, 179
- Double crossovers, 117, 118
- Double digests, and restriction mapping, 379–80
- Double helix, structure of DNA, 10–11, 206–7, 216–18
- Double-strand break model, of recombination, 347–49, 365
- Double-strand break repair, 344
- Double-stranded DNA, 221, 225
- Double transformation, 157
- Douglas, L., 30
- Down, John Langdon, 192
- Downstream direction, of DNA transcription, 249
- Downstream promoter element (DPE), 262
- Down syndrome, *177*, 192–94, *546*
- Dreyer, W. J., 494
- *Drosphila ananassae*, 182–84
- *Drosophila melanogaster* (fruit fly)
  - allelism, 354
  - chromosomal breaks, 200
  - chromosomal map, 121, *122*, *169–71*
  - chromosome number, *50*
  - chromosome puffs, 450
  - complementational analysis, 319, *320*
  - DDT resistance, 533–34, *535*, 540
  - development, 469–79
  - diagram of chromosome 2, *9*
  - disruptive selection, 578, *579*
  - dosage compensation, 94–95
  - duplications of chromosomal segments, 185–86
  - epistasis, 33
  - gene mapping, 396
  - generation interval, *66*
  - genic balance, 84–85
  - geotactic response, 541, *542*
  - giant banded chromosomes, 449
  - giant salivary gland chromosome, 119–20, 179, *439*, *455*
  - gynandromorph, *192*
  - heterozygosity, *599*
  - homeobox genes, *482*
  - infective particles, 522
  - inversion, *180*, *181*
  - linkage groups of, 110
  - as model organism, 8, 65
  - mutants and mutations, *24*
  - nomenclature, 23–24
  - nondisjunction, 190, *191*
  - nuclease-hypersensitive sites of DNA, 446
  - nucleolus, 260
  - population genetics, 566–67
  - realized heritability, 543
  - satellite DNA, *452*
  - scanning electron micrograph of, *109*
  - sex determination, 87
  - sex-linked inheritance, 95–96, 103, 104
  - size of chromosome, 441–42
  - testcrossing of, 110–11, 114–16, 118–20, 140–41
  - twin spots, *132*
  - unequal crossing over, *188*
- *Drosophila pseudoobscura*, 596, *599*
- Drug resistance. *See also* Antibiotics and antibiotic resistance
  - bacteria and, 153
  - fungi and, 124
- Drug sensitivity, 154
- Duffy blood group, 134
- Duplications, of chromosomal segments, 185–86
- Dusts, as carcinogens, *492*
- Dyad chromatid pair, 59
- Dynein, 52
- Dysplasia, 491

**E**

- Ectoderm, 470
- Edman method, 284–85
- Edward syndrome, 194, *195*
- Ehrman, Lee, 597
- Eicher, Eva, 86
- Eichwald, Ernst, 85
- Either-or rule, 72
- Electrophoresis
  - evolution and empirical data, 14, 553
  - experimental methods, 92–94
  - glucose-6-phosphate dehydrogenase system, 91
  - polymorphisms, 14, 598
- Electroporation, 374
- Ellis, J., 303
- Elongation, and translation, 288, 292–93, 296
- Elongation complex, 262–63, *264*
- Elongation factors (EF-Ts and EF-Tu), 292, *293*
- Eldredge, N., 594
- Encephalopathies, and animal diseases, 213
- Endoderm, 470
- Endogenote, 159
- Endomitosis, 119–20
- Endonucleases, 226, *334*
- Endosomic vesicles, 500
- Endotoxins, 398
- Energy cost
  - of protein biosynthesis, 313
  - of translation, 297
- Energy industry, and biotechnology, 398
- Enhancers, and transcription, 263
- *Enhancer of zeste* gene, 483
- Enriched medium, 150
- *Env* (envelope) gene, 503
- Environment
  - carcinogens, 492
  - evolutionary role of sexual reproduction, 88
  - induced versus inherited traits, 509
  - phenotypic distributions, 533, *534*
- Environmentally induced puffs, 450
- Enzymes
  - active site, 206, *208*
  - classical genetics, 10
  - DNA and control of, 218–19
  - DNA repair in *E. coli*, *348*
  - DNA replication in *E. coli*, *231*
  - Tay-Sachs disease, 23
  - three-dimensional structure, 206–7
- Enzymology, of DNA replication, 225–38
- *Eosinea* gene, 35–36
- *Ephestia kühniella* (flour moth), 510, *511*
- Epidemiology, of AIDS, 502–3
- Epistasis, 32–37, 40–41
- Epstein, R. H., 296
- *Epuliscium fishelsoni* (bacteria), 441
- Equational division, 60
- Erythromycin, *153*, 515
- Escherich, Theodor, 149
- *Escherichia coli*. *See also* Bacteria
  - autoradiography of DNA, *223*
  - conditional-lethal mutants, 328
  - conjugation, 157, *158*
  - DNA repair, *348*
  - DNA replication in, *231*, *233*, 238
  - DNA-RNA hybridization, 246
  - F factor, 161
  - fine-structure mapping, 321
  - gene expression, 406–12, 435–36
  - generation interval, *66*, 149
  - genetic variation, 316
  - heat shock proteins, 430
  - host-range mutations, 163
  - *lacZ* gene and directed mutations, 339
  - medium for cultivation, *150*
  - as model organism, 8
  - nucleotide excision repair, 341
  - N-formyl methionine and protein synthesis, 288
  - nucleolus, 278
  - phage λ, *420*
  - pili, 159
  - plasmids and cloning, *357*, 361
  - promoter of ribosomal RNA gene, *rrnB*, *249*
  - recombinant DNA techniques, 13
  - ribosome, *257*
  - RNA polymerase, *250*
  - scanning electron micrograph, *148*
  - T4 bacteriophage, *405*
  - transduction experiments, 168, 172
  - transformation, 155
  - tRNAs for methionine, *289*
- *Essay on the Principle of Population, An* (Malthus, 1798), 589
- Ethics, and genetics
  - cloning, 374–75
  - evolution, 590–91
  - human behavioral genetics, 547
  - human genome sequencing, 397
  - Lysenko affair, 6–7
  - recombinant DNA technology, 370–71
  - sociobiology, 606
- Euchromatin, 48, 452
- Eugenics, 547
- Eukaryotes
  - chromosome, 440–61, 462
  - clones and cloning, 369–70
  - control of transcription, 465–69
  - daughter cells, 49

Eukaryotes—*Cont.*
  DNA replication, 231–32, 238–39
  DNA transcription, 260–75, 278
  foreign DNA, 372–75
  gene expression, 12
  initiation factor, 290
  life cycle, 64
  nucleotide excision repair, 343
  prokaryotes compared to, 47, 261–62, *298*, 440
  size of cells, 441
Eukaryotic vectors, 369–70
Euploidy, 197–98
Eusocial hymenoptera, 606
*Even-skipped* gene, 476, 477
Evolution
  Darwinian, 5, 589, 590–91
  DNA synthesis, 228–29
  genetic code, 311–12
  hemoglobin genes, 460
  homeotic control, 479
  imprinting, 524
  intron function, 272–74
  inversions, 182–84
  mutations, 339
  population genetics, 553
  proto-oncogenes, 491
  reproductive isolating mechanisms, 607
  sexual reproduction, 88–89
  sociobiology, 603–6
  speciation, 589–94, *595*
  supergenes, 181
  transposons, 429–30
Evolutionary genetics, *4*, 5, 13–14
Evolutionary rates, 601, *603*
Ewing's sarcoma, *485*
Excisionase, 418
Excision repair, 340–46
Exogenote, 159
Exons, 265
Exon shuffling, 272–74
Exonucleases, 226–27
Experimental design, 74
Experimental methods
  adaptive mutations, 339
  amino acid sequencing, 284–86
  chromosomal painting, 486
  computer program for allelic equilibrium under heterozygote advantage, 583–84
  gene sequencing, 388–89
  high-speed chromosomal sorting, 444–45
  inversions and evolutionary sequences, 182–84
  *in vitro* site-directed mutagenesis, 333–35
  lethal equivalents, 562
  mapping of quantitative trait loci, 537–38
  mimicry, 604
  polymerase collisions, 252
  protein motifs of DNA recognition, 480–81
  size of cells, 441
  transcription in real time, 250–51
  viroids and introns, 272–73
Expression vectors, 369
Expressivity, of phenotype, 97, 98
Extrachromosomal inheritance, 509
Eye color, and concordance, *546*

**F**

Facioscapular muscular dystrophy, 582
Factorials, 72
Familial Down syndrome, 193–94
Familial insomnia, 213
Family tree, 98–99
Fate map, 471
F cells, 160, 161
F-duction, 161
Fecundity selection, 577
Feedback inhibition, and posttranslational control, 433
Felsenfeld, Gary, 221
Female-lethal gene, 85
Fertility factor (F factor), 158–59
Fetus, human
  hemoglobin, 460
  inbreeding and death rates, 562
  Rh locus and maternal incompatibility, 582
*F* factor, 161, *162*
Field mouse. *See Peromyscus polionotus*
Filial generation, 18
Fine-structure mapping, 320–23
Fingerprints, heritability of human, 544, *545*. *See also* DNA fingerprints
First-division segregation (FDS), 128–29, *130*
Fisher, Ronald A., 4, 13, 553, 604
Fitness, and natural selection, 577, 585, 605–6. *See also* Survival of the fittest
5-bromouracil (5BU), 331
5′ untranslated region (5′ UTR), 291
Flemming, W., 3
Floral induction, 479, 482
Floral meristem, 479
Floral-meristem identity genes, 483
Flow cytometry, *445*
Flowers. *See also* Plants
  genetic control of development, 479, 482–83
  inheritance of color, 23, *37*
  sex determination, 87, 90
Flow karyotype, 444
Fluctuation test, 316–17, *318*
Fluorescent in situ hybridization (FISH), 486
*FMR-1* gene, 187–88
Fokker-Planck equation, 575
Follicle cells, 470, *474*
Follicular lymphoma, *485*
Food processing, and biotechnology, 398
Footprinting, 248
Ford, Edmund B., 596
Forensics, and DNA fingerprint, 381
Four o'clock plants (*Mirabilis jalapa*), 23, 517
Founder effect, 576
Four-stranded DNA, 221
F-pili, 159
Fraenkel-Conrat, H., 210–11
Fragile site, of chromosome, 490
Fragile-X syndrome, 186–88
Fragment length polymorphisms (RFLPs), 537
Frameshift, and genetic code, 304–5
Frameshift mutation, 326–27, *338*
Franklin, Rosalind, 205, *215*
*FrdC* gene, 389
Free energy of activation, 205
Frequency-dependent selection, 597
Fruit fly. *See Drosophila melanogaster*
Functional alleles, 318, *320*
Fundamental number (NF), 185
Fungi
  haploid mapping, 123
  heterokaryon test, 509
  phenotypes, 124
  size of cells, 441
Furberg, S., 206
*Fushi tarazu* gene, 477
Futch, David, 182–84

**G**

*Gag* (group antigen gene), 503
β-Galatosidase, 406, *407*
β-Galatoside acetyltransferase, 406
β-Galatoside permease, 406
Galápagos Islands, 594
Gallo, Robert C., 502
Galton, Francis, 535, 547
Gametes
  combinations of chromosomes in, 67
  meiosis, 55
  rule of independent assortment, 27
  rule of segregation, 18
Gametic selection, 577
Gametophyte, 64
Gancylovir, 376
*Gap* genes, 475
Garden pea (*Pisum sativum*)
  chromosome number, *50*
  classical genetics, 9
  independent assortment, *27*
  Mendel's experiments, 17–18, 65
  phenotypic and genotypic ratios of dihybrid, *32*
Garrod, A. E., 37
Gastrulation, 470
G-banded chromosomes, 134, *138–39*, 451
Gehring, Walter J., 478
Gene. *See also* Gene expression; Gene mapping; Gene order
  classical genetics, 9
  HIV and, 503–4
  segregation, 18
Gene amplification, 459
Gene cloning. *See* Clones and cloning; Recombinant DNA technology
Gene conversion, 351
Gene expression. *See also* Transcription; Translation
  cancer, 484–92
  catabolite repression, 412–13
  immunogenetics, 492–501
  inducible system, 406–12, 435
  lytic and lysogenic cycles in phage λ, 418–24
  operon model, 12, 406, 408, 418
  patterns in development, 469–84
  posttranslational control, 433
  transcription factors, 430
  translational control, 430–33
  transposable genetic elements, 425–30
  *trp* operon, 413–18

Gene family, 458–59
Gene flow, 592
Gene mapping. *See also* Mapping
  human mitochondrial chromosome, *512*
  phage λ, *419*
  sequencing of human genome, 390–97
Gene order
  haploid mapping, 129–32
  Hfr strains, *162*
  three-point cross, 118
  two-factor cotransduction, *167*
Gene overlap, 388–89
Gene pool, 555
General interval, 66
Generalized transduction, 166
General transcription factors, 262, *263*
Generation time, 556–57
Gene therapy, 397
Genetically modified crops (GM crops), 371
Genetic code
  breaking of, 305
  DNA control of enzymes, 218–19
  evolution and, 311–12
  overlap and punctuation of, 305, *306*
  synthetic codons, 306–7
  synthetic messenger RNAs, 306
  triplet nature of, 304–5
  universality of, 308–11
  wobble hypothesis, 307–8
Genetic diseases
  gene therapy, 397
  hemoglobins, 460–61
  imprinting, 524
  inbreeding, 562
  mitochondrial inheritance, 514–15
Genetic drift, 555, 575–76, 585
Genetic engineering. *See* Recombinant DNA technology
Genetic fine structure, 317–24
Genetic load, 597
Genetic polymorphism, 596
Genetics. *See also* Biomedical applications; Cytogenetics; Ethics; Experimental methods; Heredity; Inheritance; Molecular genetics
  bacteria and bacterial viruses and research in, 149–50
  classical, evolutionary, and molecular, 4–5, 9–14
  definition of, 3
  fruit flies and colon bacteria as subjects of experiments, 7–8
  history of, 3–4
  Nobel laureates, *113–14*
  scientific method, 5–7
  techniques of study, 8
Genetic variation, 316, 596–603
Genic balance, in *Drosophila*, 84–85
Genome
  electrophoresis and sampling of, 598
  HIV-1, *504*
  mapping, 110
  mitochondrial, 512
  sequencing of human, 390–97
Genomic equivalence, 469
Genomic library, 366, *367*
Genomics. *See also* Recombinant DNA technology
  benefits of cloning, 397–98
  DNA sequencing, 383–90
  polymerase chain reaction, 381–83
  probing for cloned gene, 368–77
  prokaryotic vectors, 360–66
  restriction endonucleases, 359–60
  restriction mapping, 377–81
  southern blotting, 366–68
Genophore, 154
Genotypes
  evolutionary genetics, 14
  genotypic interactions, 30–37
  independent assortment, *27*
  pedigree analysis, 97–98
  segregation, 18, 21
  transposition, 429–30
Genotypic interactions, 30–37
*Geochelone elephantopus vanderburgh*. *See* Giant land tortoise
*Geospiza fortis* (Ground finch), 590–91, *595*
Geotactic response, 541, *542*
Gerald, Park, 197
Gerstmann-Straussler-Scheinker syndrome, 213
*Giant* gene, 476
Giant land tortoise (*Geochelone elephantopus vanderburgh*), *529*
Giant salivary gland chromosomes, 119–20, 179, *439*, *455*
Giemsa stain, 451
Gilbert, Walter, 272–73, 274, 383, 390, 412
β-Globin gene, 367–68
Globin gene family, 459–61
Glucose-6-phosphate dehydrogenase (G-6-PD), 90–91
Glycolytic pathway, *10*
Goodfellow, Peter, 86
*Gooseberry* gene, 477
Gould, Stephen J., 594
Grandfather method, and X linkage, 132–33
Grant, B. & P., 590
Grasshopper (*Chorthippus parallelus*), *59*
Green fluorescent protein, 377
Griffith, F., 154, 209
Ground finches, of Galápagos Islands, *588*, 590–91, 594, *595*
Ground squirrels, 606
Group I and Group II introns, 266
Group selection, 603–4, 606
Grunberg-Manago, M., 306
*Gryllus domesticus*. *See* Cricket
G-tetraplex, 456, *457*
Guanine, 213, *215*
Guide RNA (gRNA), 275
*Gurken* gene, 474–75
Gynandromorphs, 191, *192*

## H

*Haemanthus katherinae*, 53, *54*
*Haemophilus influenzae*, 396
Hair color, and concordance, *546*
Hairless mutation, of *Drosophila*, 24
Haldane, J. B. S., 4, 13, 553
Hall, B., 246
Hamilton, W. D., 605–6
Hammerhead ribozyme, 266, *270*
Hamster, chromosome of, *448*
Handedness (left or right), and concordance, *546*
Haploid cells, 9, 48
Haploidiploidy, 606
Haploid mapping, 122–32
Hard selection, 597
Hardy, G. H., 13, 553
Hardy-Weinberg equilbrium
  alleleic frequencies, 553–54
  definition of, 553
  extensions of, 558–59
  generation time, 556–57
  migration, 573
  multiple loci, 559–60
  mutation, 572, 573
  natural selection, 555
  proof of, 555–56
  random mating, 554–55, 560
  selection models, 578
  small population size, 574
  testing for fit of, 557–58
HAT medium, 135
Heart disease, congenital, *545*
Heat, and transversions, 336
Heat shock proteins, 249–50, 430
Helicase II, 342
Helix-turn-helix motif, 480
Helper T cells, 502
Hemizygous genes, 96
Hemoglobin
  human development, 459–61
  sickle-cell anemia, 39
Hemophilia, 95, 100–102
Heredity, chromosomal theory of, 66. *See also* Genetics; Inheritance
Heritability. *See also* Inheritance
  measurement of, 544–45, 548
  partitioning of variance, 543–44
  realized, 542–43
Hermaphrodites, 191
Hershey, A. D., 209–10
Hertwig, O., 3
Heterochromatin, 48
Heteroduplex DNA, 351, *352*
Heterogametic chromosomes, 83
Heterogeneous nuclear mRNAs (hnRNAs), 265
Heterokaryon test, 134, 509
Heteromorphic chromosome pair, 49, 83
Heteroplasmy, 511
Heterothallic mating type, 469
Heterotrophs, 150
Heterozygote advantage, 582, 583–84, 596–97
Heterozygotes, 21, 22–23
Heterozygosity, 564–65, *599*
Heterozygous DNA, 351
Hexoseaminidase-A, 23
Hfr strains, 158–59, 160, 161, *162*
High-speed chromosomal sorting, 444–45
*Hind*II restriction enzyme, 360
Histocompatibility Y antigen (H-Y antigen), 85

Histone acetyl transferases (HATs), 446
Histone genes, *459*
Histone proteins, 443
Histones, composition of, *447*
History, of genetics, 3–4, 17, 28–29, 30–31, 47
HIV, and genes, 503–4. *See also* AIDS
Ho, David, 504
Hogness box, 262
Holandric traits, 95
Holland, J. J., 276
Holliday, R., 347
Holliday junction, 347, *350*
Holoenzyme, 231
Homeo box, 478–79, *482*, 505
Homeo domain, 478, *479*, 505
Homeotic genes, 477–78
Homogametic chromosomes, 83
Homogentisic acid, 38
Homologous chromosomes, 48
Homologous recombination, 347, 375
Homology-directed recombination, 344
Homomorphic chromosome pairs, 48, 83
Homoplasmy, 511
*Homo sapiens*. *See* Humans
Homothallic mating type, 469
Homozygosity, and inbreeding, 561–62, 562, 564
Homozygotes, 21, 22–23
Horseshoe crab. *See Limulus polyphemus*
Host-range mutations, 163
Hot spots, and mutations, 323
Hox clusters, 479
Hubby, J. L., 596, 598
Human Genome Project, 6, 390–97, 537
Humans (*Homo sapiens*). *See also* Fetus
  aneuploidy, 192–97
  behavioral genetics, 547
  chromosomal maps, 132–40
  chromosomal rearrangements, 186–90
  chromosome number, *50*
  composition of DNA, *216*
  fingerprints and heritability, 544, *545*
  flow karyotype of chromosomes, *444*
  gene mapping and sequencing of genome, 390–97
  generation interval, *66*
  hemoglobins, 459–61
  heterozygosity, *599*
  karyotype, *49*
  linkage groups of, 110
  melanin synthesis, *36*
  metacentric, submetacentric, and acrocentric chromosomes, *49*
  mitochondria and mitochondrial inheritance, 512, 514–15
  monosomy, 200
  nondisjunction of sex chromosomes, *191*
  nuclease-hypersensitive sites of DNA, 446
  quantitative inheritance, 545–46
  Rh locus and maternal-fetal incompatibility, 582
  sex determination, 83, *84*, 85–86, *87*
  telomeric sequence, *455*
  xeroderma pigmentosum, 343
Humoral immunity, 492
*Hunchback* gene, 475, 476
Hungry codons, 432
Huntington disease, 188, 582
Hybrid DNA, and recombinant DNA technology, 351
Hybridomas, 494
Hybrid plasmid, 360
Hybrids and hybridization, 18, 199
Hybrid vectors, 361, 363
Hybrid vehicle, 360
Hybrid zones, 592
Hydrogen bonding, in DNA, 218
Hyperplasia, 491
Hypervariable loci, 380–81
Hypothesis testing, 74–76
Hypotrichosis, *100*
Hypoxanthine phosphoribosyl transferase (HPRT), 134–35

**I**

Identity by descent, 561
Idiogram, 48, *49*
Idiotypic variation, 493
Idling reaction, 432
*Ijij* genotype, 517–18
Imino acid, 281
Immune system, and adenosine deaminase, 397
Immunity, definition of, 492
Immunity reactions, of ABO blood types, *25*
Immunogenetics, 492–501
Immunoglobulins
  definition of, 492
  formation of, 498
  genes and antibodies, *497*
  hybridomas, 494
  variability of, 493, 506
Imprinting, 509, 524
Imprinting center (IC), 524
*Inborn Errors of Metabolism* (Garrod 1909), 37
Inbreeding
  definition of, 554
  genetic diseases, 562
  genotypic proportions in population with, *563*
  homozygosity from, 561–62
  nonrandom mating, 560
Inbreeding coefficient, 561, 567
Inbreeding depression, 542
Incestuous unions, 99
Inclusive fitness, 605–6
Incomplete dominance, 23
Independent assortment
  classical genetics, 10
  example of, *34*
  meiosis and mitosis, *63*, 67
  rule of, 26–30
Independent events, 71–72
Indian fern (*Ophioglossum reticulatum*), *50*
Induced mutation, 325–26
Inducer, and gene expression, 408
Inducible system, of gene expression, 406–12, 435
Induction, and gene expression, 165, 409, 489–91
Industry, and genetic engineering, 398
Infective particles, and cytoplasmic inheritance, 518–22
Information transfer, and translation, 281–303
Inheritance. *See also* Cytoplasmic inheritance; Genetics; Heredity; Heritability; Quantitative inheritance
  determination of non-mendelian, 509
  heritability, 542–45
  polygenic inheritance, 533–35, 541
  population statistics, 535–40
  selection experiments, 541–42
  traits controlled by multiple loci, 531–35
Initiation codon, 288
Initiation complex, and translation, 288–91
Initiation factors (IF1, IF2, IF3), 289
Initiation signals, for transcription, 248
Initiator Element (Inr), 262
Initiator proteins, 230
Insecticides, 398
Insertion mutagenesis, 488
Insertion sequences (IS elements), 425, *426*
Insertions, and mutations, 336–37
Insomnia, familial, 213
Instantaneous speciation, 593
Interbreeding, 591
Intercalary heterochromatin, 453
Intergenic suppression, and mutations, 337–38
Integrase, 418
Intelligence quotient (IQ), and heritability, 546
Interkinesis, 59
Internal ribosome entry site, 291
Interphase, of mitosis, 48, *53*, *56*, *62*
Interpolar microtubules, 53
Interrupted mating, 159–61, *162*
Intersex, 85
Intervening sequences. *See* Introns
Intestinal bacterium. *See Escherichia coli*
Intra-allelic complementation, 323–24
Intragenic suppression, 327
Introns
  eukaryotic transcription, 265, *266*
  evolution, 272–74
  self-splicing of, *271*
  single-stranded DNA loops, *268*
  *Tetrahymena* ribozyme, *269*
  viroids, 272–73
Introns-early and introns-late views, of exon shuffling, 273, 274
Inversion, and chromosome breaks, 179–81, 200
Inverted-repeat sequence, 251
*In vitro* research, 209
*In vitro* site-directed mutagenesis, 333–35
*Iojap* chromosomal locus, 516
Ionizing radiation, *492*
Iron oxide, *492*
IS elements, 425, *426*, 427
Isochromosome, 194, *195*
Isopropyl oil, *492*
Isozymes, 94

**J**

Jacob, François, 159–60, 161, 409
Jeffreys, A., 381
Johannsen, Wilhelm, 21, 541
Junctional diversity, 495
Junk DNA, 458–59

**K**

Kaposi's sarcoma, 502
Kappa particles, 520–21, *522*
Karotype, 48, *49*
Karpechenko, G. D., 198
Karyokinesis, 47
Kavenoff, Ruth, 440–41
Khruschev, Nikita, 7
Killer paramecium, 520–21, 525
Kimura, Motoo, 562, 599
Kinesin, 52
Kinetochore, 48, *55*, 453
Kinetochore microtubules, 53
Kinetoplasts, 275
King, Mary-Claire, 391
Kin selection, 605–6
Kleckner, Nancy, 429, *430*
Klenow fragment, 232
Klinefelter syndrome, 85, 197
Klotz, L., 440–41
Klug, A., 480
*Knirps* gene, 476
Knockout mice, 375–76
Koehn, R., 599
Kornberg, Arthur, 225
*Krüppel* gene, 476
Kuhn, Thomas, 591
Kuru, 213
Kynurenin, *39*, 510, *511*

**L**

*Lac* operon, and gene expression, 406–12, 436
Lactate dehydrogenase (LDH), 93, *94*
Lactose metabolism, 406, *407*, 436
*LacZ* gene, 339
Ladder gels, 386, 399
Lagging strand, 225
Lahn, Bruce, 89
Lamarck, Jean-Baptiste, 5
Lampbrush chromosomes, 450–51
Landsteiner, Karl, 25
*Lathyrus odorata*. *See* Sweet pea
Leader, and RNA transcription, 255
Leader peptide gene, 416–19
Leader transcript, 415–16
Leading strand, 225
Leber optic atrophy, 514–15
Leder, Phillip, *307*
Lederberg, Joshua, 157, 158, 161, 165, 166
*Leishmania tarentolae*, *275*
Lejeune, Jérôme, 192–93
Lemmings, and population density, 597
Leptonema, 56
Lethal-equivalent alleles, 561, 562
Leucine zipper, 480, *481*
Leukemia, 484, *485*
Level of significance, 77–78
Lewin, B., 523
Lewis, Edward B., 478
Lewontin, Richard C., 596, 598
LexA protein, 346, *347*
Life cycle
  of bacteriophages, 163–65
  of eukaryotes, 64
  generalized for animals and plants, *47*
  of *Paramecium*, 518–19
  of plants, 64–66
  of yeast, *126*
Ligation, and DNA replication, 227–29
*Liguus fasciatus* (Tree snail), *570*
Lily (*Lilium longiflorum*), *50*
*Limenitis archippus*. *See* Viceroy butterfly
*Limnea peregra* (Pond snail), *510*
*Limulus polyphemus* (Horseshoe crab), *599*
Linear measurement, metric units of, *48*
Linkage
  classical studies of, 66
  equilibrium and disequilibrium, 559–60
Linkage groups, 110
Linkage number, 236
Linkers, and blunt-end ligation, 363–64, *365*
Liposarcoma, *485*
Liposome-mediated transfection, 374
Liposomes, 374
Literature, scientific, 6–7
Littlefield, J. W., 134
Liver, and sickle-cell anemia, 39
Location, of DNA, 208–9, 219
Lock-and-key model, of enzyme functioning, 206–7
*Lod* score method, 135
Long interspersed elements (LINES), 458
Lovell-Badge, Robin, 86
Luciferase gene, 377
Luciferin, 377
Luft disease, 514
Luria, Salvador E., 4, 316, 317
Lwoff, André, 423
Lymphocytic cell, and mitosis, *15*
Lymphomas, 484
Lyon, Mary, 90
Lyon hypothesis, 90–91
Lysate, 163, 164
Lysenko, T. D., 6–7
Lysogenic cycles, in phage λ, 418–24
Lysogeny, 165
Lytic cycles, in phage λ, 418–24

**M**

MacLeod, C., 209
Major histocompatibility complex (MHC), 492
Malaria, 596
Male-specific lethal (*msl*) gene, 94
Malthus, Thomas, 589
Maniatis, Tom, 461
Manic-depression, and concordance, *546*
Map distances, 116–18, 131
Mapping. *See also* Chromosomal maps; Gene mapping; Restriction mapping
  classical studies of, 66
  conjugation, 161
  diploid, 110–11, 114–22
  haploid, 122–32
  human chromosomal maps, 132–40
  quantitative trait loci, 537–38
  sequencing of human genome, 390–97
  transduction, 166–68
  transformation, 155–57
Mapping function, 117–18
Map units, 10, 111
Margulis, Lynn, 513
Mariculture, and biotechnology, 398
Mass production methods, and Human Genome Project, 396
*Master-switch* genes, 478, 505
Mate-killer infection, 521–22
Maternal effects
  definition of, 509
  maternal-effect genes, 472, *473*, 475
  moth pigmentation, 510, *511*
  snail coiling, 509–10
Maternal-effect genes, 472, *473*, 475
Mating types, 124, 469
Matthei, J. H., 306
Maturation-promoting factor (MPF), 50
Maxam, Allan, 383
McCarthy, B. J., 276
McCarty, M., 209
McClintock, Barbara, 120–22, *123*, 425, 468
McKusick, Victor A., 99, 137
Meadow weed. *See Arabidopsis thaliana*
Mean, statistical, 74, 535, 536, 538–39
Mean fitness, of population, 579–80
Measles, and concordance, *546*
Measurement
  heritability, 544–45, 548
  metric units of linear, *48*
Medicine, and genetic engineering, 397. *See also* Biomedical applications
Meiosis, *62*
  animals and, 63–64
  definition of, 47
  origins of term, 49
  process of, 55–61
  rules of segregation and independent assortment, 67
  significance of, 61, 63
  triploids, *197*
Meiosis II, 60–61
Meiotic drive, 577
Melanin, 33, 34, *36*
Melanoma, *485*
Mendel, Gregor
  classical genetics, 9–10
  dominance, 22–23
  experiments in genetics, 17–18, 97
  genotypic interactions, 30–37
  history of genetics, 3, 4, 65
  independent assortment, 26–30
  multiple alleles, 25–26
  nomenclature, 23–24
  quantitative traits, 534
  rediscovery of, 47
  rule of segregation, 18–22, 63
  statistics, 71, 74
  techniques of study, 8
Merozygote, 159, 409–10
Meselson, Matthew, 220–21
Mesoderm, 470

Messenger RNA (mRNA)
  cloning, 364–65, *366*
  function of, 245
  polycistronic and monocistronic, 298, 299, *300*
  prokaryotes, 277
  synthetic, 306, *307*
Messing, J., 386
Metabolic pathways, of color production in dihybrids, *36*. *See also* Metabolism
Metabolism, and biochemical genetics, 37–38
Metacentric chromosome, 48, *49*
Metafemales, 85
Metagon, 521
Metamales, 85
Metaphase, of mitosis, 52, 54, *57*
Metaphase I, and meiosis, 59, *60*, *62*
Metaphase plate, 54
Metastasis, 484
Methionine, *152*, 288, *289*
Methylation, of DNA, 466–67
Metrical variation, 533
Metric measurements, linear, *48*
MHC proteins, 498–500, *501*
Michigan Technological University, 441
Microsatellite DNA, 382
Microscope, history of, 3
Microtubule organizing centers, 52
Microtubules, 48, 52
Midparent, and wing length, *539*
Migration
  Hardy-Weinberg equilibrium, 555
  population genetics, 573–74
Mimicry, 604, *605*
Minimal medium, 124, 150
*Mirabilis jalapa*. *See* Four o'clock plants
Misalignment mutagenesis, 337
Mismatch repair, 343–44, 354
Missense mutations, 338
Mitochondrial DNA (mtDNA), 600
Mitochondrial transfer RNAs, 309
Mitochondrion
  cytoplasmic inheritance, 511–15
  genetic code of yeast, *311*
  scanning electron microscope image of, *508*
  signal hypothesis, 303
Mitosis
  definition of, 47
  late anaphase, *56*
  lymphocytic cell, *15*
  origins of term, 49
  process of, 52–55
  rules of segregation and independent assortment, 67
  significance of, 55
  stages of, *46*, *57*
Mitosis-promoting factor, 50
Mitotic crossing over, 132
Mitotic spindle, 52–53, *55*
Mixed families, of codons, 307
Model organisms, 8
Modern linkage map, 393, 395
Molecular chaperones, 303
Molecular evolutionary clock, 600–602
Molecular genetics. *See also* Cytogenetics
  Down syndrome, 193
  evolutionary processes, 553
  as general area of study in genetics, *4*, 5, 10–13
  growth in field of, 358
  search for genetic material, 205–11
Molecular imprinting, 524
Molecular mimicry, 297
Moloney murine leukemia virus, 374
Monarch butterfly (*Danaus plexippus*), 604, *605*
Monocistronic messenger RNAs, 299
Monoclonal antibody, 493
Monod, Jacques, 409
Monoecious flowers, 87
Monohybrids, 18
Monosomic cells, 190
Monosomy, 195, 200
Monovalent chromosome, 59
Monozygotic twins, 546
Montagnier, Luc, 502
Morgan, Thomas Hunt, 83, 95, 110, 111
Morphogen, 472
Morphological species concept, 591
Morton, Newton E., 135
Mosaicism, 90, 190–92, 194
Moths, and pigmentation, 510, *511*
Mouse (*Mus musculus*). *See also* Deer mouse
  allelic frequency, *600*
  c-banding of chromosomes, *452*
  chromosome number, *50*
  epistasis and color of, 33–34, *35*, *37*
  generation interval, *66*
  genetic bottlenecks, 577
  heterozygosity, *599*
  knockout, 375–76
  nuclease-hypersensitive sites of DNA, 446
  population density, 597
  transgenic, 373
  Turner syndrome, 196
Müller, F., 604
Muller, H. J., 88, 121, 325–26, 330
Müller-Hill, B., 412
Müllerian mimicry, 604, *605*
Muller's ratchet, 88
Multihybrids, 30, *31*
Multilocus control, of inheritance, 532–33
Multilocus selection models, 598–99
Multinomial expansion, 73, 558
Multiple alleles, 25–26
Multiple loci, 532–33, 559–60
Mu particles, 521–22
Muscular dystrophy, 13, 188
Mustard gas, *492*
Mutability, of DNA, 207
Mutational equilibrium, 571–73
Mutation rates, 326, 572–73
Mutations and mutants
  adaptive, 339
  AIDS virus, 504
  amino acids, 313
  *bicoid* gene, 472
  cancer, 484–87, 506
  chemical mutagenesis, 330–31, 336–37
  classical genetics, 10
  colinearity, 324–25
  definition of, 316
  evolutionary theory, 5, 88, 312
  fluctuation test, 316–17
  genetic fine structure, 317–24
  Hardy-WEinberg equilibrium, 555
  homeotic genes, 477–78
  intergenic suppression, 337–38
  *in vitro* site-directed mutagenesis, 333–35
  *lac* operon, 409–12
  misalignment mutagenesis, 337
  mutator and antimutator, 338
  nomenclature, 24
  point mutations, 326–28
  population genetics, 571–73
  selection-mutation equilibrium, 581
  sexual reproduction, 88
  spontaneous mutagenesis, 329–30
  spontaneous versus induced, 325–26
Mutator mutations, 338
Muton, 320
Mutually exclusive events, 71
*Mycobacterium tuberculosis*, *216*, *546*
Mycoplasmas, 441

**N**

Nail-patella syndrome, 134, 135
*Nanos* gene, 472–73, 474
Nathans, D., 4, 359
National Academy of Sciences, 370
National Institutes of Health
  gene therapy, 397
  guidelines on recombinant DNA technology, 370
  Human Genome Research Institute, 358
Natural selection
  effects of, 577–78
  evolution, 5, 14
  fitness, 577
  Hardy-Weinberg equilibrium, 555
  models of, 581–82, 584
  recessive homozygote, 578–81
  selection-mutation equilibrium, 581
*Nef* gene, 504
Negative assortative mating, 554
Negative interference, 118
N-end rule, 433
Neo-Darwinism, 13, 589
Neomycin, 375, 376
Neoplasms, 484
*Neurospora crassa* (Bread mold)
  biosynthesis of niacin, 38–39
  chromosome number, 49
  generation interval, *66*
  haploid mapping and, 123, 141
  mutations of, 124
  ordered spores, 125–27
  phenylalanine synthesis, 41
Neutral alleles, 599
Neutral gene hypothesis, 598
Neutral petite, 514
Newt. *See Notophthalmus viridescens*; *Triturus*
N-formyl methionine, 288
Niacin, 38, 39
Nickel, as carcinogen, *492*
Nicotindamide adenine dinucleotide (NADH), 93
Nilsson-Ehle, H., 532

Nirenberg, Marshall W., 305, 306
Nitrous acid, 331
*Nivea* gene, 35–36
Nobel, Alfred, 112
Nobel Prize, in genetics, 112–14
Nomenclature, 23–24
Noncoding strand, of DNA, 249
Nondisjunction
  mosaicism, 190–92
  sex chromosomes, 83
Nonhistone proteins, 448
Non-Hodgkin's lymphoma, *485*
Nonhomologous chromosomes, 182–85
Nonhomologous end joining, 344
Non-Mendelian inheritance, 509
Nonparental ditype (NPD), 125, *127*
Nonparental phenotypes, 111
Nonrandom mating, 560–65
Nonreciprocity, and sex linkage, 96
Nonrecombinant phenotypes, 111, 118
Nonsense codons, 296–97
Nonsense mutations, 337–38
Normal distributions, 535, *536*, *538*
Northern blotting, 368
*Notophthalmus viridescens* (newt), *451*
Novitski, E., 30
N segments, 496–97
Nuclease-hypersensitive sites, 445–46
Nucleic acids, chemistry of, 211–19
Nucleolar organizer, 53
Nucleolus, 53, 260–61, 278
Nucleoprotein, 47, 440, 442–61
Nucleoside, 212, *214*
Nucleosome structure, 442–47
Nucleotides, 8
  biomedical applications of triple-stranded chains, 221
  conserved sequence, 277–78
  DNA replication, *226*
  excision repair, 341–44
  nomenclature, *215*
  nucleic acids, 211–12, 213–14
  techniques of genetic research, 8
Null hypothesis, 77–78
Nullisomic cells, 190
Numerator elements, 85
Nurse cells, 470
Nüsslein-Volhard, Christiane, 471
Nutritional-requirement mutants, 327
Nutritional requirements
  of bacteria, 151, 153
  of fungi, 124

## O

Ochoa, Severo, 305, 306
Okazaki fragments, 225, *228*, *229*, *230*
Oligonucleotide primer, 387
Oncogenes, 13, 484, 488–92
*One-gene-one-enzyme* rule, 10, 38–39
Onion (*Allium cepa*), *46*
Oocyte, nucleus and follicle cells, *474*
Oogenesis, 64
Oogonia, 64
Open reading frames (ORFs), 291
Operator, and gene expression, 408–9, 414
Operon model, 12, 406, 408, 418
*Ophioglossum reticulatum*. *See* Indian fern
Ordered spores, 125–27
Original literature, 6–7
Origin recognition complex (ORC), 239
*Origin of Species, The* (Darwin), 3, 589
*Ostreococcus tauri* (green alga), 441
Outbreeding, 554, 560
Ovary determining (Od) gene, 86
Overlap, of genetic code, 305, *306*
Ovum, 64
Oxidative phosphorylation, 512
*Oxytricha nova*, 456, *457*

## P

Pachynema, 56
Page, David, 85, 89
*Pair-rule* genes, 475
Palindrome, 359, *361*
Palmer, J., 274
Panmictic population, 574
Paracentric inversion, 180
Paramecin, 520
*Paramecium*, 518–19, 525
Parameters, statistical, 538
Parapatric speciation, 592, *593*
Parasegments, 470, *471*
Parental ditype (PD), 125, *127*
Parental imprinting, 524
Parental phenotypes, 111
Parthenogenesis, 64
Partial digest, 379
Partial dominance, 23
Partitioning, of variance, 543–44
Pascal's triangle, 73
Patau syndrome, 194–95
Path diagram, 562–64, *565*
Pattern formation, and flower development, 479, 482
Pauling, Linus, 206, 214–15, 602
pBR322 cloning plasmid, 363, *364*, *372*
Pea. *See* Garden pea; Sweet pea
Pearson, K., 535
Pedigree analysis, 97–102, 562–64, *565*, 567
Peer review, and scientific journals, 7
Penetrance, of genetic trait, 97–98, 103, 104
Penicillin, 153, 154, 172, 294
Pepsin, 284
Peptic ulcer, *545*
Peptide bond formation, 292
Peptide fingerprint, 284
Peptide map, 284
Peptidyl site (P site), 292
Peptidyl transferase, 292
Pericentric inversion, 180
Permissive temperature, and mutation, 327
*Peromyscus maniculatus*. *See* Deer mouse
*Peromyscus polionotus* (Field mouse), *599*
PEST hypothesis, 433
Petal whorl, 479, *482*
Petite mutations, 514, 525
Petroleum, as carcinogen, *492*
P53 gene, 487, *488*
  Phage, and cloning with restriction enzymes, 361. *See also* Bacteriophages
Phage labeling, 209–10
Phage M13, *390*
Phage λ, lytic and lysogenic cycles in, 418–24, 435–36
Phage resistance, 163
Phage T4, 430
*Phaseolus vulgaris*. *See* Bean
Phenocopy, 98
Phenotypes
  of bacteria, 151–54
  blood groups, 25–26
  *Drosophila melanogaster*, 8
  of fungi, 124
  independent assortment, 29
  *lac* operon, 436
  segregation, 21
  transposition, 429–30
  of viruses, 154
Phenylalanine, 41
Phenylisothiocyanate (PITC), 285
Phenylketonuria (PKU), 558, 582
Pheromones, 469
Phosphodiester bonding, 214
Phosphoglycerate kinase, *283*
Photocrosslinking, 250
Photoreactivation, 340
Phyletic evolution, 590
Phyletic gradualism, 594, *595*
Phylogenetic tree, 601
Physical crossover, 120–22
Physical map, and Human Genome Project, 393, 395
Pigeon (*Columba livia*), *50*, *59*
Pigment and pigmentation. *See also* Color; Skin color
  epistasis in mice, 33–34
  maternal effects in moths, 510, *511*
Pili (fimbriae), 159
Pink bread mold. *See Neurospora crassa*
Pistil, 87
*Pisum sativum*. *See* Garden pea
Plants. *See also Arabidopsis thaliana*; Flowers; Garden pea; *Zea mays*
  genetically altered crops, 398
  genetic control of development, 479, 482–83
  life cycle, *47*, 64–66
  polyploidy, 198–99
  proplastid formation and variegation, 516
  sex determination, 87, 90
  vectors and cloning of, 371–72
Plaques, bacterial, 151
Plasmids
  bacteria in genetic research, 149
  *E. coli* and, *357*
  *F* factor, 158
  prokaryotic and cytoplasmic inheritance, 522–24
  recombinant DNA technology, 13
Plastids, 515
Pleiotropy, 39
Plus-and-minus method, of DNA sequencing, 383, 388
*Pneumocystis carinii*, 502
Pneumonia, and AIDS, 502
Point centromere, 453
Point mutations, 326–28
Polar body, 64
Polarity, of DNA structure, 218

Polar mutants, 425
Pole cells, 470
*Pol* (polymerase) gene, 503
Politics, and ethics in genetics, 6–7
Pollen grain, 65
Poly-A tail, 265, *267*
Polycistronic messenger RNAs, 297, *300*
Polydactyly, 99, *100*
Polygenes and polygenic inheritance, 533–35, 541
Polymerase chain reaction (PCR), 381–83
Polymerase collisions, 252
Polymerase cycling, 231
Polymerization, of nucleotides, 214, *215*
Polymorphisms, 14, 596–99. *See also* Variation
Polynucleotide phosphorylase, 306
Polyploids, 83, 197–98
Polyribosome, 297
Polysome, 297, *300*
Polytene chromosomes, *439*, 449, *455*
Pond snail. *See Limnea peregra*
Population, definition of, 553
Population analysis, 564–65
Population density, 597
Population genetics
  computer programs to simulate, 583–84
  evolution, 553, 589
  Hardy-Weinberg equilibrium, 553–60
  migration, 573–74
  models for, 571
  mutation, 571–73
  natural selection, 577–84
  nonrandom mating, 560–65
  small population size, 574–77
Population statistics, 535–40
*Populus tremuloides* (Quaking aspen), 441
Position effect, 179
Positive assortative mating, 554
Positive interference, 118
Postreplicative repair, of DNA, 344–46
Posttranscriptional modifications, 261–62
Posttranslational control, 433
Posttranslocational state, of ribosome, 293, *298*
Postzygotic mechanisms, 592
Potato spindle tuber viroid (PSTV), 272–73
Poultry, and realized heritability, *543*. *See also* Chicken
PPL Therapeutics, 374
Prader-Willi syndrome, 524
Preemptor stem, 417
Preer, J., 520, 521
Preinitiation complex (PIC), 262
Pretranslocational state, of ribosome, 293, *298*
Prezygotic mechanisms, 592
Pribnow box, 248
Primary oocytes, 64
Primary spermatocytes, 63
Primary structure, of protein, 281
Primary transcript, 265
Primase, 225–26
Primers
  creation of general-purpose, 386–87, 390
  DNA replication, 225–27, *228*, 232
Prions, 213
Probability, 71–74, 78
Probability theory, 71
Proband, 98
Product rule, 72
Progeny testing, 21–22
Prokaryotes. *See also* Bacteria; *Escherichia coli*
  DNA transcription, 246–55, 277, *465*
  eukaryotes compared to, 47, 261–62, *298*, 440
  genomics, 360–66
  mitochondrial ribosomal RNA, 512–13
  operon model and, 12
  origins of term, 149
  plasmids, 522–24
  ribosomes, *301*
Prokaryotic vectors, 360–66
Prolactin, and signal peptide, *302*
Promoters
  efficiency of, 430
  *lac* operon, *408*
  repressor operon of phage λ, 435–36
  transcription in eukaryotes, 262–65
  transcription in prokaryotes, 248–51
Proofreading, and DNA polymerase, 227
Prophage, 165
Prophase, of mitosis, 52, 53, *57*
Prophase I and II, and meiosis, 56–60, *62*
Proplastids, 515, 516
Proposita and propositus, 98
Proteasome, 499
Protein degradation, and posttranslational control, 433
Protein electrophresis, 92–94
Protein-folding problem, 303
Protein-mediated splicing, 268–72
Proteins
  DNA and control of, 205–7
  DNA repair in *E. coli*, *348*
  energy requirements of biosynthesis, 313
  methods of depiction, *283*
  motifs of DNA recognition, 480–81
  primary, secondary, tertiary, and quaternary structures, 281
  synthesis of, *245*, *283*, 287
Proteomics, 396–97
Proto-oncogenes, 13, 489, 491
Prototrophs, 150
Prusiner, Stanley, 213
Pseudoalleles, 319
Pseudoautosomal regions, 87, 95
Pseudodominance, 96, 178
Ptashne, Mark, 412
Punctated equilibrium, 14, 594, *595*
Punctuation, of genetic code, 305, *306*
Punnett, Reginald C., 21, 26
Punnett square diagram, *21*, 26, 28, 30, 532
Purine nucleotides, 213, *214*
Puromycin, 294, *295*
Pyrimidine nucleotides, 213, *214*

**Q**

Quantitative inheritance, 533, 534, 537–38, 545–46. *See also* Inheritance
Quaternary structure, of proteins, 281

**R**

Rabl, K., 3
Racism, and sociobiology, 606
RAG recombinase, 495, *496*
Random genetic drift, 555, 575–76, 585
Random mating, 554–55, 556
Random strand analysis, 122
*Raphanobrassica* (cabbage-radish cross), 198
*Raphanus sativus* (radish), 198
*Ras* oncogene, 491
R-bands, 452
Read-through process, and terminators, 251
Realized heritability, 542–43
Real time, and observation of transcription, 250–51
RecA protein, 345–46, 435
RecBCD protein, 349–51, 354
Recessive homozygote, and natural selection, 578–81
Recessive inheritance and recessive traits, 18, 99–100, 102
Reciprocal cross, 18, 115
Reciprocal translocation, 182, *186*, *187*
Recombinant DNA technology. *See also* Genomics
  bacteria, 349–51
  bacteriophages, 163–64
  double-strand break model of, 347–49
  ethical debate on, 370–71, 374–75
  genome sequencing, 110
  heteroduplex DNA, 351, *352*
  hybrid DNA, 351
  meiosis, 61, 63
  overview of techniques, 358
  restriction endonucleases, 12–13
Recombinant phenotypes, 111, 116
Recombinant plasmids, 360, *362*
Recombination nodules, 59
Recombination signal sequences, 494
Recon, 320
Red clover, 26
Red eye, mutation of *Drosophila*, 24
Reductional division, 59
Redundant controls, and amino acid operons, 418
Regional centromeres, 453
Regression and regression analysis, 539–40
Regulator gene, 406
Reintiation, of translation, 291
Relaxes mutant, 432
Release factors (RF), 296
Repair, of DNA
  categories of systems, 339–40
  damage reversal, 340
  in *E. coli*, *348*
  excision repair, 340–46
Repetitive DNA, 458
Replication, of DNA
  control of enzymes, 219
  enzymology of, 225–38
  in eukaryotes, 231–32, 238–39
  mutability, 207
  process of, 220–24
  structures and, 238
Replication-coupling assembly factor, 444

Replicons, of DNA, 224
Replisome model, of DNA replication, 234, *235*
Reporter systems, 376–77, 399
Repressible system, 406
Repression, gene expression and maintenance of, 421–23
Repressor, and gene expression, 406, *408*, *409*
Repressor transcription, 421
Reproductive isolating mechanisms, 592, 607
Reproductive success, 577
Repulsion, of alleles, 111
Rescue experiment, 472
Residues, of amino acids, 281
Resistance transfer factor (RTF), 523
Resolution, of transposition, 428
Resolvase, 428
Restricted transduction, 165
Restriction digest, 378
Restriction endonucleases, 12–13, 359–60
Restriction enzymes, 360–61, 400
Restriction fragment length polymorphisms (RFLPs), 378, 380–81
Restriction mapping, 377–81
Restriction sites, 359
Restrictive temperature, and mutation, 327
Retinoblastoma, 486–87
Retrotransposons, 458
Retroviruses
  AIDS, 502–4
  cancer, 487–88
Reverse bands, 452
Reverse transcription, 276
Reverse transcriptionase, 276
Reversion, and mutation rates, 326
*Rev* (regulation of expression of virion proteins) gene, 504
Rhesus system, and blood types, *32*
Rh locus, and maternal-fetal incompatibility, 582
Rhoades, M., 516
Rho-dependent terminators, 251, 254, *255*
Rho-independent terminators, 251, *255*
*Rhomboid* gene, 466
Rho protein, 251
Ribosomal RNA (rRNA), 245, 246
Ribosome recylcing factor (RRF), 297
Ribosomes
  messenger RNA, 245
  transcription, 256
  translation, 290–91, 292, 299, 301
Ribozyme, 266
Rice, and genetic modified varieties, 398
Rice, William, 89
Rich, Alexander, 219, 221
Rickets, vitamin-D-resistant, 97–98, 102, 509
Ritland, D., 604
RNA. *See also* Messenger RNA; Ribosomal RNA; Transfer RNA
  antisense, 431
  computer model of serine transfer, *243*
  editing, 275
  genetic code dictionary of, *12*
  genetic material and, 210–11
  guide, 275
  priming of DNA synthesis, 228, *230*, 232
  self-replication, 276
  types of, 245–46
  viruses, 213
RNA phages, 276
RNA polymerase
  observation of, *250*
  prokaryotic and eukaryotic, 247, *260*
  transcription, 11, 12
RNA replicase, 276
RNA tumor viruses, 276
Roberts, Richard J., 265, *266*
Robertson, F., 544
Robertson, W., 185
Robertsonian fusion, 185
Rodriguez, R., 363
Rolling-circle model, of DNA replication, 238, *239*
Roslin Institute, 374
Roundworm. *See Ascaris* spp.
Rous, Peyton, 487
R plasmids, 522–23
RRE (rev response element) gene, 504
*R* II screening techniques, 321, *322*, *323*
R222 plasmid, 523
  Rule of independent assortment, 27–29, 67
Rule of segregation, 18, 21–22, 63, 67
*Rumex acetosa*. *See* Sorrel

**S**

*Saccharomyces cerevisiae* (Baker's yeast)
  centromeres, 453
  genetically modified, 398
  haploid mapping, 123
  telomeres, 456
  unordered spores, 124–25
Sager, Ruth, 517
*Salmonella typhimurium*
  Ames test for carcinogens, 332
  generalized transduction, 166
  transposon orientation, 429, 468
Sampling distribution, 74
Sampling error, 574–75
Sanger, Frederick, 284, 383
Sarcomas, 484
Satellite DNA, 451, *452*, 458
Scaffold structure, of chromatin, 448, *449*
Scanning hypothesis, 291
Schizophrenia, *545*, *546*
*Schizosaccharomyces pombe* (fission yeast), 111
Scientific journals, 7
Scientific method, 5–7, 71
Scrapie, 213
Screening, for drug resistance, 154
Sea urchin, *216*
Secondary oocyte, 64
Secondary sources, and scientific literature, 7
Secondary spermatocytes, 64
Secondary structure, of proteins, 281
Second-division segregation (SDS), 128–29, *130*, 131
Second messenger, 412
Securin, 54
Segmentation genes, 475–77
*Segment-polarity* genes, 475
Segregation
  classical genetics, 10
  meiosis and mitosis, 67
  rule of, 18–22, 63, 67
  translocation, 183–85
Segregational load, 597
Segregational petite, 514
Segregation distortion, 577
Selection coefficient, 577
Selection experiments, 541–42
Selection model, 578–79
Selection-mutation equilibrium, 581
Selective medium, 150
Selenocysteine, 310–11
Selfed. *See* Self-fertilization
Self-fertilization, 17, 21–22, *31*
Selfish DNA, 430
Self-splicing, 265–66, *268*
Semiconservative replication, of DNA, 220
Semisterility, 181
Sendai virus, 134
Sepal whorl, 479, *482*
Separin, 54
Sequence-tagged sites (STSs), 393
*Sequoiadendron giganteum* (giant redwood), 441
Seryl tRNA synthetase, *280*
*Serratia marcescens* (bacteria), *151*
Sex chromosomes, 83, *191*, 197
Sex-conditioned traits, 96
Sex determination, 83–90
Sex-determining region Y (SRY), 86
Sexduction, 161
Sex-influenced traits, 96
Sexism, and sociobiology, 606
Sex-lethal (*Sxl*) gene, 85, 95
Sex-limited traits, 96
Sex linkage, 95–96, 104
Sex-linked inheritance, 100–102, 103, 104
Sex-ratio phenotype, 522
Sex-reversed individuals, 85–86
Sex-switch gene, 85
Sexual reproduction, and evolution, 88–89
Sexual selection, 577
Shapiro, J. A., 427–28
Sharp, Philip A., 265, *266*
Sheep (*Ovus aries*)
  cloning of, *1*, 374–75
  generation interval, *66*
  realized heritability, *543*
Shepherd's purse (*Capsella bursa-pastoris*), *37*
Sheppard, P. M., 604
Sherman, P., 606
Sherman Paradox, 187
*Shigella*, and resistance plasmids, 523
Shine-Dalgarno hypothesis, 290–91, *300*
Shoot apical meristem, 479
Short interspersed elements (SINES), 458–59
Shotgun cloning, 364, *367*
Shub, D., 274
Shunting, and translation initiation, 291
Sickle-cell anemia, 39, 460–61, 596, *597*
Sigma factor, 247

Signal hypothesis, 301–3
Signal peptidase, 301
Signal peptide, 301, *302*
Signal recognition particle (SRP), 301
Signal-sequence receptors, 302
Signal sequences, and mitochondrial genomes, 512
Signal transduction, 467–69
Signal transduction pathway, 467
*Silene latifolia*. *See* White campion
Simian immune deficient viruses (SIVs), 502
Singer, B., 211
Single breaks, 178
Single-nucleotide polymorphisms (SNPs), 393
Single-strand binding proteins, 232, 234
Sister chromatids, 52
Site-specific recombination, 418
Site-specific variation, in codon reading, 309–10
Size
  of cells and chromosomes, 441–42
  of population, 555, 574–77
Skin color, and quantitative inheritance in humans, 545, *546*
Skolnick, M., 391
Sleeping sickness, 275
Small nuclear ribonucleoprotein particles (snoRNPs), 261, 268, *270*
Small nucleolar RNAs (snoRNAs), 261
Small population size, 574–77
SMC proteins, 444
Smith, H., 4, 359
Smoking, and cancer, *492*
Snails, and direction of coiling, 509–10, 525. *See also* Pond snail; Tree snails
Snapdragon. *See Antirrhinum majus*
Sobrero, Ascanio, 112
Sociobiology, 14, 603–6
*Sociobiology: The New Synthesis* (Wilson, 1975), 603
Soft selection, 597
Solenoid model, for chromatin fiber, *448*
Somatic-cell hybridization, 134–40
Somatic crossing over, 132
Somatic doubling, 198
Somatic hypermutation, 497
Sonneborn, Tracy M., 519, 520
Sorrel (*Rumex acetosa*), 90
SOS box, 346
SOS response, 346
Southern, Edwin, 368
Southern blotting, 366–68
Soviet Union, and Lysenko affair, 6–7
Spätzle protein, *467*
Specialized transduction, 165
Speciation, and evolution, 589–94, *595*
Species
  allopolyploidy and cross-fertilization between, 198
  definition of, 553
  inversion process and evolutionary history of groups, 181
  speciation, 589–94, *595*
Specific transcription factors, 263
Speckles, and nuclear messenger RNAs, 268
Spermatids, 64
Spermatogenesis, 63–64
Spermatogonium, 63
Sperm cells, 63
Spiegelman, S., 246
Spina bifida, and concordance, *546*
Spindle pole body, 52
Spiral cleavage, 510
Spirillum, 149
Spliceosome, 268–72
Splicing, of exons, 265
Splicing factors, and introns, 271
Spontaneous mutagenesis, 328–30
Spontaneous mutation, 325–26
Sporophyte, 64
Stability, of mutational equilibrium, 571–72
Stabilizing selection, 578
Stadler, L. J., 325–26
Stage-specific puffs, 449–50
Stahl, Franklin W., 220–21
Stalin, Joseph, 6
Stamen, 87
Stamen whorl, 479, *482*
Stanbridge, E., 487
Standard deviation, 74, 536, 538–39
Standard error of the mean (SE), 539
Standard method, and Human Genome Project, 392
Stanford University, 371
Starvation, and attenuator regulation, 417
Statistics and statistical methods, 74–78
Stature, and heritability of human traits, *545*
Steiz, Joan A., 268, *269*
Stem-loop structure, 252
Stepladder gels, 386, 387
Stern, Curt, 132
Stochastic events, 71
*Streptococcus pneumoniae*, 209
*Streptomyces antibioticus*, *151*
*Streptomyces griseus*, *151*
Streptomycin, *153*, 294, 517–18
Stringent factor, 432
Stringent response, 432
Strobel, S., 221
Strong inference, and scientific method, 5
Structural alleles, 318–19, *320*
Sturtevant, Alfred H., 4, 121, 186
Submetacentric chromosome, 48, *49*
Substrate transition, and DNA replication, 329
Subtelocentric chromosome, 48
Sucrose density-gradient centrifugation, 256
Sulfur bacteria, 441
Summer squash (*Cucurbita pepo*), *37*, 40–41
Sum rule, 72
Supercoiling, 234, 236, *237*
Supergenes, 181
Superinfection, and repression, 422
Superrepressed mutation, 412
Suppressive petite, 514
Suppressor gene, 337
Suppressor of Stellae (*Su* or *Ste*) gene, 87
Surveillance mechanisms, and cell cycle, 51
Survival of the fittest, 589. *See also* Fitness
Sutton, Walter, 4, 66, 83, 110
SV40 virus, 371, *372*, 378
Svedberg, T., 256
Svedberg unit (S), 256
Sweet pea (*Lathyrus odorata*), *37*
Swine, and realized heritability, *543*
SWI/SNF complex, 446–47
Sympatric speciation, 592–94
Synapsis, and meiosis, 56
Synaptonemal complex, 58
Syncitium, 470
Syn-expression group, 465
Synteny test, 136
Synthetic codons, 306–7
Synthetic medium, 150
Synthetic messenger RNAs, 306, *307*
Szostak, J., 347

**T**

TACTAAC box, 268
Tanksley, Steven, 538
*Tat* (trans-activating transcription factor), 503–4
TATA-binding protein (TBP), 262
TATA box, 262, *263*
Talmud, 95
TAR (trans-activating response element), 504
Tatum, Edward L., 10, 38–39, 157, 161
Tautomeric shifts, 328, *329*
Taylor, J., 440
Tay-Sachs disease, 23, 582
T/B-cell lymphoma, *485*
TBP-associated factors (TAFs), 262
T-cell receptors, 492, 498, *499*
Telocentric chromosome, 48
Telomerase, 454–55, 456–57
Telomeres, 178, 454–57
Telophase, of mitosis, 52, 54–55, *56*, *57*
Telophase I, of meiosis, 59–60, *62*
Telson, 471
Temin, Howard, 276
Temperate phages, 165
Temperature-sensitive mutants, 327
Template, and DNA replication, 220
Template strand, of DNA, 249
Template transition, and DNA replication, 329
Termination
  of DNA replication, 236, *238*
  of translation, 296–301
Termination signals, for transcription, 248
Terminators, and transcription, 248, 251–55
Terminator sequence, 248
Terminator stem, 416
Terramycin, *153*
Tertiary structure, of proteins, 281
Testcross and testcrossing. *See also* Three-point cross; Two-point cross
  dihybrids, *33*
  examples of, 140–41
  multihybrids, 30
  segregation rule, 22
Testing, for AIDS, 504
Testis-determining factor (TDF), 85
Tetracycline, *153*, 294
Tetrad analysis, and haploid mapping, 122–32
Tetrads, 59
*Tetrahymena thermophila*, *269*, 272–73

Tetranucleotide hypothesis, 216
Tetraploids, 83, 197
Tetratype (TT), 125, *127*
Thalassemia, 460, 461, 582
*Thermus aquaticus* (hot-spring bacteria), 382
Theta structure, of DNA, 223–24
*Thiomargarita namibiensis* (sulfur bacteria), 441
3-hydroxyanthranilic acid, 38, 39
Three-locus control, 532
Three-point cross, 114–16, 118–20
Thymidine kinase (TK), 134–35
Thymine, 213, *215*
Time frame, for population equilibrium, 580–81
Tissue-specific puffs, 450
T-loop, and telomeres, 456, *458*
T-lymphocytes, *464*
Tobacco mosaic virus, 210–11, *212*
Tobacco plants, and recombinant DNA technology, 377
*Toll* gene, 473
Topoisomers and topoisomerases, 236, *237*
*Torpedo* gene, 474–75
*Torso* gene, 473
Tortoise. *See Geochelone elephantopus vanderburgh*
Totipotent cells, 469
Touchette, Nancy, 213
Trailer, and RNA transcription, 255
Trans-acting mutations, 411
Trans configuration, 111, 133
Transcription. *See also* Gene expression
  classical genetics, 11–12
  control of in eukaryotes, 465–69
  definition of, 244
  early and late in phage infection of *E. coli*, 420–21
  eukaryotic DNA, 260–75
  flow of genetic information and, 244, 275–76
  observation of in real time, 250–51
  prokaryotic DNA, 246–55
  prokaryotic initiation and termination signals for, 248
  transfer RNA, 256–60
  translation, 297, 299
  types of RNA, 245–46
Transcription factors, 430, 466–69
Transducing particle, 166
Transduction, 154, 165–68, 172
Transfection, 372
Transfer operon (tra), 523
Transfer RNA (tRNA)
  function of, 245
  transcription, 256–60
  translation, 281–83, 286–88, 292
Transfer RNA loops, 258, 260
Transformation
  bacteria and bacteriophages, 154–57
  DNA and, 155, *156*, 209
  in eukaryotes, 371
Transformation mapping, 155–57, 172
Transformed cancers, 484
Transgenic organisms, 372, 397
Transient polymorphism, 597
Transition mutation, 329, 331, 336
Translation. *See also* Gene expression
  classical genetics, 12
  control of, 430–33
  definition of, 244
  DNA involvement, 276
  energy cost of, 297
  genetic code, 304–12
  information transfer, 281–303
  initiation of, 288–91
  start and stop signals of, 313
  transcription, 297, 299
Translational control, 430–33
Translocation
  segregation, 183–85
  translation and elongation, 292–93, 296
Translocation channel (translocon), 301
Transposable genetic elements, 425–30
Transposase, 428
Transposition, mechanism of, 427–28
Transposons, 425–30, 468, 469
Transversion mutation, 329–30, 336
Tree snail. *See Liguus fasciatus*
Trihybrids, 30, *33*, 40
Triplet nature, of genetic code, 304–5
Triplex DNA, 221, *222*
Triple-X female, 197
Triploids, 83, 197
Trisomic cells, 190
Trisomy 8, 195
Trisomy 13, 194–95
Trisomy 18, 194, *195*
Trisomy 21, 192–94
*Triticum* wheat, 198
*Triturus* (newt), *260*
*Trp* operon, 413–18
*Trp* RNA-binding attenuation protein (TRAP), 417–18
True heritability, 543
Trypanosomes, and RNA editing, 275
Trypsin, 284
Tryptophan, 39, 413–14, 416–17
Tuberculosis, *216*, *546*
Tumor-inducing (*Ti*) plasmid, 371
Tumors, and cancer, 484
Tumor-suppressor genes, 485–87
Turner syndrome, 85, 195–96
Twin spots, and mitotic recombination, 132
Twin studies, of concordance, 546
*Twist* gene, 466
2-aminopurine (2AP), 331
Two-dimensional chromatography, 284–86
Two-locus control, 531–32
Two-point cross, 110–11, 114
Type I and type II errors, 75
Typological thinking, 590
Tyrosine, *36*
Tyrosine kinases, 491

## U

UAA, UAG, and UGA stop codons, 309
Ubiquitin, 51
Ultraviolet (UV) light
  as carcinogen, *492*
  dimerization, 340
Unequal crossing over, 186
Uninemic chromosome, 440, *442*
Unique DNA, 458
Universality, of genetic code, 308–11
University of California, 371
Universisty of Toronto, 441
Unmixed families, of codons, 307
Unordered events, 72
Unordered spores, and yeast, 124–25
Unusual bases, of transfer RNA, 257–58
Upregulation, of cell growth, 487
Upstream direction, of DNA transcription, 249
Upstream element (UP element), 249
Uracil, 213, *215*, 344
Uracil-DNA glycosylase, *342*
U-tube experiment, *158*
*UvrC* and *uvrD* genes, 341, 342

## V

Van Breden, E., 3
Vancomycin, 294
Van Leeuwenhoek, Anton, 3
Variable-number-of-tandem-repeats (VNTR), 381
Variance
  partitioning of, 543–44
  statistical methods, 536, 538–39
Variation. *See also* Polymorphisms
  additive models, 533, 538, 548
  causes of, 316
  chromosome number, 190–99
  codon-reading and site-specific, 309–10
  continuous, 17, 531, 533
  Darwinian evolution, 313, 589, 596
  discontinuous, 17, 531
  frequency-dependent selection, 597
  heterozygote advantage, 596–97
  multilocus selection models, 598–600
  patterns of, 600–603
Variegation, of color, 179
Vavilov, Nikolai, 6
V(D)J joining, 496
Venter, J. Craig, 358, 396
Viceroy butterfly (*Limenitis archippus*), 604, *605*
*Vicia faba*. *See* Broad bean
Victoria (Queen of England), 100–102
*Vif* gene, 504
Vinyl chloride, *492*
Virchow, Rudolf, 3
Virion, 154
Viroids, and introns, 272–73
Viruses. *See also* Bacterial viruses; Retroviruses
  cancer, 13, 487–92
  cultivation of, 151
  life cycle, 64
  operon model, 12
  phenotypes of, 154
  recombinant DNA technology, 371
  RNA, 213
Vitamin-D-resistant rickets, 97–98, 102
V-J joining, 494, 495, *497*
V-*myc* gene product, 491

Von Mohl, Hugo, 3
Von Nageli, C., 3, 48
Vonnegut, Kurt, 213
Von Tschermak, Erich, 4, 17
*Vpr* gene, 504
*Vpu* gene, 504
Vries, Hugo de, 4, 17

## W

Wahlund effect, 574
Waldeyer, W., 3, 48
Wallace, Douglas, 514, 597
Walzer, Stanley, 197
Washburn, Linda, 86
Watson, James, 4, 10, 11, 205, 215, 216, 328, 330
Weiling, F., 31
Weinberg, W., 13, 553
Weismann, August, 3
Western blotting, 368, 369
Wheat, and grain color, 531–33. *See also Triticum* wheat
White campion (*Silene latifolia*), 90
Whole-genome shotgun method, 392
Wieschaus, Eric F., 471
Wild-type
  metabolic pathways in *Neurospora*, 41
  phenotypes of *Drosophila*, 23–24
Wilkins, Maurice, 205
Williams, G., 604–5
Williams, R., 210–11
Wilm's tumor, 487
Wilson, Edward O., 603
Wobble hypothesis, 307–8, 309, *310*, 312
Wollman, E., 159–60, 161
Wood and leather dust, *492*
Wright, Sewall, 4, 13, 553
Wynne-Edwards, V. C., 603–4

## X

X chromosome, 90–95, 100
Xeroderma pigmentosum, 343, 484
X-gal, 386–87
X inactivation center (XIC), 91
X linkage, 95–96, 132–33
XO-XX system, of sex determination, 86
X-ray crystallography, of DNA, 215–16
XY system, of sex chromosomes, 83, 87, 90
XYY karyotype, 196–97

## Y

Yanofsky, Charles, 324, 415
Y chromosome, 87, 88–89
Yeast. *See also Saccharomyces cerevisiae*; *Schizosaccharomyces pombe*
  antibiotic resistance, 515
  centromeres, *454*
  computer model of transcriptional factor, *244*
  genetic code and mitochondria, *311*
  mating type, 469
  mitochondrial DNA, 513, 515
  nucleotide excision repair, 343
  petite mutations, 514, 525
  testcrossing, 111, 141
  transfer RNA, *259*
  unordered spores, 124–25
  vectors and cloning, 370
Yeast artificial chromosomes (YACs), 370
Y-junctions, of DNA, 223, 232, 234, *236*
Y-linked chromosomes, 95
Yunis, J., 486–87

## Z

Z DNA, 219, *220*, 467
*Zea mays* (corn)
  Ac-Ds system, 468
  cytoplasmic inheritance, 516–17
  epistasis, 34–35, *37*
  generation interval, *66*
  genotypic interactions and phenotypes, 32
  life cycle of, *65*, 66
  meiosis, 60, *62*
  testcross of trihybrid, 40
Zeigler, D., 252
ZFY gene, 86
Zimm, B., 440
Zinc finger, 480, *481*
Zinder, N. D., 166
Zuckerkandl, E., 602
ZW system, 87
Zygonema, 56
Zygotes, and gene pool concept of formation, *555*
Zygotic induction, 165
Zygotic selection, 577